

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G06F 17/30	A1	(11) International Publication Number: WO 00/51024 (43) International Publication Date: 31 August 2000 (31.08.00)
(21) International Application Number: PCT/IL00/00117 (22) International Filing Date: 25 February 2000 (25.02.00) (30) Priority Data: 60/121,596 25 February 1999 (25.02.99) US (71) Applicant (for all designated States except US): TEXTRAY LTD. [IL/IL]; Kehilat Venezia Street 12, 69400 Tel Aviv (IL). (72) Inventors; and (75) Inventors/Applicants (for US only): DAGAN, Ido [IL/IL]; Shivtei Israel Street 26a, 47267 Ramat Hasharon (IL). STAUBER, Yitzhak [IL/IL]; Remez Street 32, 53242 Givatayim (IL). (74) Agent: REINHOLD COHN AND PARTNERS; P.O. Box 4060, 61040 Tel Aviv (IL).		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: METHOD AND APPARATUS FOR DYNAMICALLY DISPLAYING A SET OF DOCUMENTS ORGANIZED BY A HIERARCHY OF INDEXING CONCEPTS		
(57) Abstract <p>A system that includes a processor associated with a memory and display for dynamically presenting set of documents to users. The memory is configured to store a predetermined hierarchy of indexing concepts and is configured to store a set of documents. The processor is configured to provide hierarchical display of the indexing concepts. The indexing concepts are associated with the set of documents. The processor is further configured to apply the following as many times as required: determining a subset of documents by utilizing the hierarchical display thereby rendering it organized document subset, defining at least one indexing concept in the hierarchical display so as to constitute an "organizing" concept, and providing an organizing hierarchical display of indexing concepts, wherein the root of the organizing hierarchical display is the organizing concept, wherein concepts in the organizing hierarchical display are associated with the organized document subset.</p>		

PTO/SB/08A Ref. No.: F3

Serial No.: 09/815,591

Docket No.: DOC-002/01US

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD AND APPARATUS FOR DYNAMICALLY DISPLAYING A SET OF DOCUMENTS ORGANIZED BY A HIERARCHY OF INDEXING CONCEPTS

5

FIELD AND BACKGROUND OF THE INVENTION

The amount of textual information that is available in computerized media has increased dramatically in recent years. As a result, there is an increasing need
10 for end users to have effective tools for searching, browsing, navigating, reading and analyzing collections of textual documents. Current common practice, within organizations as well as in the Internet, is having a search engine that indexes a large repository of documents and enables users to issue a search query and to get in response all documents that satisfy the search conditions. Usually, a list of
15 titles, along with some additional information, is presented for each document and the user can further ask for the display of specific documents from the list. The list of documents is often sorted by some relevance ranking, which is intended to approximate the degree of relevance of the document to the query. Sorting by date is also often available.

20 A search mechanism typically attaches to each document a set of *indexing concepts*. An indexing concept is a symbol or value that characterizes the document, and is typically used within search queries or within routing queries ("queries" that specify which documents will be routed to an addressee). Typical types of indexing concepts include (but are not limited to):

- 25 1. Topical categories (also known as controlled keywords, topics, descriptors etc.). These are symbols denoting topical issues, which are usually general or abstract concepts that do not necessarily appear

literally in the text. For example, a topical category may be "Company Acquisition". This term, serving as the name of the category, may not appear literally in a document that describes such an event.

2. Important terms and names of entities (such as countries, companies, products and people) which appear or are referred to in the text (as is or by synonyms).
3. Document meta-data items, such as document source, type, author and date.

In the following, a document is considered *indexed* by the indexing concepts characterizing it. Apart from being used in ad-hoc search queries, indexing concepts may also be used to determine routine routing of incoming documents to addressees.

The process of associating indexing concepts to documents (the *indexing* process) is performed either manually, automatically, or by some combination of the two modes. With respect to indexing concepts that consist of terms and names from the document text, the indexing process usually involves scanning the text of the document, identifying words, terms and names, and possibly bringing these terms to some canonical form (e.g. the grammatical base form (lemma) of the word). Meta-data indexing concepts are often determined by the systems, in which the document is created or received, but may also be handled manually.

Of particular interest to the invention is the indexing process for topical categories (*categories*, in short). In many systems, it is possible for the user to manually assign topical categories to a document. More recently, there have been developed a number of methods for assigning topical categories to documents automatically, which are referred to here as automatic text classification methods. Such methods classify documents to appropriate categories taken from a predetermined list of possible categories. Classification is performed by some mechanism that receives the document text as input and determines the appropriate categories based on the words, terms or their combinations that

appear in the document.

There are two common approaches for automatic text classification methods. The first approach is based on manual definition of the rules, or some other type of logic, by which a document is being classified to a category based on the terms in the text. For example, some systems allow users (or administrators) to define complex queries, which may include Boolean and other types of conditions (such as weights and proximity) that the terms in the document should satisfy. A document that satisfies these conditions is classified to the category. An example for such a system is the Topics TM system that was developed by Verity Inc., USA.

The second approach is based on automatic learning of the "logic" which entails the classification of the document to a category. Methods belonging to this approach utilize a set of *training* documents, for which the correct categories are known in advance (usually as the result of manual classification of these documents). A learning method may then include a learning phase, in which some model of the category is constructed. For example, such a model may include terms that are highly associated with the category, and possibly some weights that quantify the degree of correlation (entailment) between each term and the category. Alternatively, a learning method may be *memory based*, in which case the learning method simply stores the training data in some useful format. Then, when a new document is given for classification, the method classifies it automatically by consulting or applying the category model (or by simply comparing the document to the training data, in case of a memory based approach). Examples for trainable (learning) classification systems are described in:

1. C. Apte and F. Damerau and S. Weiss, 1994. Towards language independent automated learning of text categorization models, in Proceedings of ACM-SIGIR Conference on Information Retrieval.

2. W.W. Cohen, Text categorization and relational learning, in Machine Learning Journal, 1995, pages 124—132.
3. W. W. Cohen and Y. Singer, Context-sensitive learning methods for text categorization, in Proceedings of the 19th Annual Int. ACM Conference on Research and Development in Information Retrieval, 1996, pages 307—315.
4. D. Lewis, 1992, An evaluation of phrasal and clustered representations on a text categorization problem, in Proc. of the 15th Int. ACM-SIGIR Conference on Information Retrieval, pages 37—50.
5. D. Lewis and M. Ringuette, 1994, A comparison of two learning algorithms for text categorization, in Proc. of Symposium on Document Analysis and Information Retrieval, pages 81—93.
6. D. Lewis and R. E. Schapire and J. P. Callan and R. Papka, 1996, Training algorithms for linear text classifiers, in SIGIR '96: Proc. of the 19th Int. Conference on Research and Development in Information Retrieval.
7. K. Tzeras and S. Hartmann, 1993, Automatic Indexing Based on Bayesian Inference Networks, in Proc. of 16th Int. ACM SIGIR Conference on Research and Development in Information Retrieval, pages 22—34.
8. E. Wiener and J. Pedersen and A. Weigend, 1995, A neural network approach to topic spotting, in Symposium on Document Analysis and Information Retrieval, pages 317—332.

Once documents have been obtained by a user, as a result of some search or some routing mechanism, these documents are typically displayed in one of

- 5 -

several formats. A common method for display is to present a list of items, each providing some high level information about a document, such as the document title, meta-data items (such as author, source or date) and possibly a short summary. The list may be sorted by document publication date or by some
5 relevance score, which quantifies the degree of relevance of the document to the user's query, as hypothesized by the search system. Another display method is a hierarchical display, in which documents are organized in a hierarchical structure, similar to a graphical user interface displaying a hierarchical file system.

U.S. patent 5,924,090 (Krellenstein) "Method and Apparatus for
10 Searching a Database of Records" discloses system for searching a database and present to the user a small number of categories along with a list of most relevant records that satisfy a query. The methodology of the Krellenstein patent has a sophisticated clustering algorithm that includes three primary steps: identifying candidate categories, weighting candidate categories and displaying a set of
15 search result categories selected from the candidate categories. A typical result of the system according to the Krellenstein patent is illustrated in Fig. 1, as extracted from the www.northernlight.com site.

Thus, as shown the query *text categorization* (1) results in 19,215 documents (records) (2) (of which 6 are shown in the first page). The documents
20 are assigned to 15 categories (3). The set of categories are determined after applying the specified sophisticated clustering including identifying candidate categories, weighting candidate categories and displaying a set of search result categories selected from the candidate categories. In accordance with the specified system, the user can repeat this process further narrowing the search
25 with each iteration. Thus, double clicking the category *Special collection documents* (4) will result in applying the specified steps again giving rise to the search results illustrate in Fig. 2. As shown there are 2057 records (5) in the sought category (6) that, in turn are assigned to 12 categories (7). As readily arises from the search results depicted in Fig. 2, the resulting categories are
30 determined dynamically and, accordingly, each search is likely to give rise to

- 6 -

different set of categories. This approach has a significant shortcoming in that every time there is a different list of categories, so the user depends on "luck" on whether the categories of interest are included in the list or not. In addition, there is no fixed structure that the user knows and can expect, in order to look for the categories that are of interest to him.

Several systems and method provide a summarization mechanism, which produces automatically a summary for a document. The summary is produced based on various rules or other criteria that evaluate the degree of importance of different parts of the document. The summary is typically constructed as an extract of important sentences or paragraphs taken from the document. For example, systems that offer summaries include the LinguistX software package from InXight Inc., USA, the "AutoSummarize" option in Word, available from Microsoft Inc., USA.

When displaying the full text of a document, many search systems highlight the search words that were matched in the document text.

The current common practice for utilizing textual information does not satisfy sufficiently the increasing need of individuals and organizations. Searching information in large repositories is often a very tedious process, preventing effective utilization of information that is potentially available to the user. In particular, searches made with current techniques in large repositories often retrieve large document sets, making it extremely difficult and often impractical for the user to browse and sift through the retrieved documents and extract the relevant knowledge hidden in the vast amount of information. The bottleneck in information quest processes thus becomes the amount of time necessary for users to satisfy their information needs, as current processes require too much of the user's time.

There is accordingly a need in the art to provide for a system and method that substantially reduces or overcomes the drawbacks of hitherto known techniques, and for increasing the effectiveness of user effort in information quest processes.

30

SUMMARY OF THE INVENTION

The invention provides for a method for dynamically presenting set of documents to users . comprising:

- 5 (a) providing a predetermined hierarchy of indexing concepts;
- (b) providing a set of documents;
- (c) providing hierarchical display of the indexing concepts such that each indexing concept in the hierarchy is associated with a sub-set of documents from among said set of documents;
- 10 (d) applying steps that include the following (i) to (iii) . as many times as required:
 - (i) determining a subset of documents by utilizing the hierarchical display, thereby rendering it organized document subset;
 - 15 (ii) defining at least one indexing concept in the hierarchical display so as to constitute a respective organizing concept; and
 - (iii) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the respective at least one organizing concept; each concept in said at least one organizing hierarchical display is associated with a respective subset of documents, from among said organized document subset.
- 20
- 25

The invention further provides for a method for presenting set of documents

to users comprising:

- (a) providing indexing concepts and a set of documents ; the set of documents are associated with concepts in accordance with triggering terms in said documents;
- 5 (b) selecting a document from said set;
- (c) selecting at least one concept associated with said document;
- (d) quantifying the importance of the triggering terms of said at least one concept and in response thereto, determining the important triggering terms; and
- 10 (e) emphasizing the important triggering terms that correspond to said at least one concept.

The invention further provides for a method for presenting set of documents to users comprising:

- 15 (a) providing indexing concepts and a set of documents ; the set of documents are associated with concepts in accordance with triggering terms in said documents;
- (b) selecting a document from said set;
- (c) selecting at least one concept associated with said document;
- (d) quantifying the importance of the triggering terms of said at
20 least one concept and in response thereto, determining the important triggering terms; and
- (e) obtaining a summary based on said important triggering terms.

Still further the invention provides for a method for dynamically presenting set of documents to users, comprising:

- 25 (a) providing a predetermined hierarchy of indexing concepts;
- (b) providing a set of documents;
- (c) providing hierarchical display of the indexing concepts; the indexing concepts are associated with the set of documents
- (d) determining a subset of documents by utilizing the

- 9 -

hierarchical display, thereby rendering it organized document subset;

(e) defining at least one indexing concept in the hierarchical display so as to constitute a respective "organizing" concept; and

(f) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the respective at least one organizing concept, wherein concepts in said organizing hierarchical display are associated with the organized document subset;

(g) repeating steps (d) to (f), as many times as required.

Still further, the invention provides for a system that includes a processor associated with a memory and display for dynamically presenting set of documents to users, comprising:

(a) the memory is configured to store of a predetermined hierarchy of indexing concepts;

(b) the memory is configured for store a set of documents

(c) the processor is configured to provide hierarchical display of the indexing concepts such that each indexing concept in the hierarchy is associated with a sub-set of documents from among said set of documents;

(d) the processor is configured to apply steps that include the following (i) to (iii), as many times as required:

(i) determining a subset of documents by utilizing the hierarchy of display, thereby rendering it organized document subset;

(ii) defining at least one indexing concept in the hierarchy of

- 10 -

display so as to constitute a respective "organizing" concept;
and

- 5 (iii) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the respective at least one organizing concept; each concept in said at least one organizing hierarchical display is associated with a respective subset of documents, from among said organized document subset.

Yet further, the invention provides for a system that includes a
10 processor associated with a memory and display for presenting a set of documents to users comprising:

- (a) the memory is configured to store indexing concepts and a set of documents ; the set of documents are associated with concepts in accordance with triggering terms in said documents;
- 15 (b) the processor is configured to select a document from said set;
- (c) the processor is configured to select at least one concept associated with said document;
- (d) the processor is configured to quantify the importance of the triggering terms of said at least one concept and in response thereto, determining
20 the important triggering terms; and
- (e) the processor is configured to emphasize in the display the important triggering terms that correspond to said at least one concept.

The invention provides for a system that includes a processor associated
25 with a memory and display for presenting set of documents to users comprising:

- (a) the memory is configured to store indexing concepts and a set of documents ; the set of documents are associated with concepts in accordance with triggering terms in said documents;
- (b) the processor is configured to select a document from said set;

- 11 -

- (c) the processor is configured to select at least one concept associated with said document;
- (d) the processor is configured to quantify the importance of the triggering terms of said at least one concept and in response thereto, determining the important triggering terms; and
- (e) the processor is configured to obtain a summary in said display based on said important triggering terms.

Still further, the invention provides for a system that includes a processor associated with a memory and display for dynamically presenting set of documents to users, comprising:

- (a) the memory is configured to store a predetermined hierarchy of indexing concepts;
- (b) the memory is configured to store a set of documents;
- (c) the processor is configured to provide hierarchical display of the indexing concepts; the indexing concepts are associated with the set of documents;

the processor is configured to apply the following steps (d) to (f) as many times as required.

- (d) determining a subset of documents by utilizing the hierarchical display, thereby rendering it organized document subset;
- (e) defining at least one indexing concept in the hierarchical display so as to constitute a respective "organizing" concept; and
- (f) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the respective at least

one organizing concept, wherein concepts in said organizing hierarchical display are associated with the organized document subset;

BRIEF DESCRIPTION OF THE DRAWINGS

5 For better understanding, the invention will now be described by way of examples only, with reference to the accompanying drawings in which:

Figure 1 – illustrates a screen result of a database search system in accordance with the prior art;

10 Figure 2 - illustrates a screen result of a database search system in accordance with the prior art;

Figure 3 – illustrates a generalized computer system.

Figure 4 – illustrates a flowchart of the preferred embodiment of the invention.

Figure 5 – illustrates a top pane the concept hierarchical display, Left Top pane
15 tree representation of hierarchical document set display, Bottom document list of a document subset.

Figure 6 – illustrates a left Top pane pie: representation of hierarchical document set display.

Figure 7 – illustrates a left Top pane: pie representation of hierarchical document
20 set display.

Figure 8 – illustrates an overlapping window - Top pane: document important terms Bottom pane: document full text and terms highlighting.

Figure 9 – illustrates a left Top pane: a document subset that have been
“organized by”. Right Top pane: the topics that have performed the
25 “organization”.

Figure 10 – illustrates an overlapping window - Top pane: document important terms. Bottom pane: document summary and terms highlighting.

- 13 -

Figure 11 – illustrates a left Top pane: tree representation of hierarchical document set display.

Figure 12 – illustrates a left Top pane: tree representation of hierarchical document set display Overlapping window - Top pane: automatic important terms selection. Bottom pane: document text and automatic selected terms highlighting.

Figure 13 – illustrates a left Top pane: a document subset that have been “organized by” twice. Right Top pane: the topics that have performed the second “organization”; and,

Figures 14 to 21 illustrate a succession of screen results obtained by applying the method in accordance with one embodiment of the invention.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

It should be noted that in the context of the invention, the terms concept and category are used interchangeably. In connection with some embodiments the term node signifies concept or category.

The invention provides novel methods for utilizing textual information that considerably increase the effectiveness of the end user when dealing with large volumes of documents. A typical embodiment of the invention is used in a computer system, as illustrated in e.g. in Fig. 3. The computer system (30) includes a processor unit (31) with input and output (32 and 33) and associated display (32) and memory (not shown). The computer system (30) is configured to display documents and information about them in order to fulfill some information needs of end users (referred in the following as “system”). The invention is, of course, not bound by any specific realization of computer system and may include any known structure such as conventional Personal Computer (P.C.) in either stand-alone or network configuration, all as required and appropriate.

- 14 -

Fig. 4 provides a high-level flow chart of a typical embodiment of the invention within some computer system (the details of the components of the invention are described below). The system presents a document set (41) in a hierarchical display (42). The structure of the display may be modified
5 dynamically by an "organize by" operation (43) maintaining, however, a predetermined structure of the hierarchy. The user may select a node (standing for indexing concept) (44) within the hierarchy, and ask for a display of information about the documents that are associated with the selected node (45). The displayed information may include one or more of the following the number of
10 sub-set of the documents that are associated with the specified indexing concept, the percentage thereof from among the entire document set, the document title, meta-data elements (such as source and date) and optionally a short summary of the document. The information is of course not limited to the specified details and may vary, depending upon the particular application.

15 The user may then select a particular document (16) for display, leading to the display of the full document text or of a summary of the document. The content of the summary, as well as highlighting within the text, are determined automatically by some indexing concepts, that are determined by default to be in focus of attention of the user. The user may then select different indexing
20 concepts to be in focus, leading to modified highlighting and summary.

The rest of the section describes the details of the preferred embodiment of the invention.

Setting and Input

25 Document set

In accordance with the invention, there is provided a method and system for presenting document sets and their content to the user of a system in an effective manner. The invention thus refers to any situation in which some *document set* has to be presented by the system, at any point of time, for purposes such as
30 exploration, scanning, reading or analysis. The term document should be

construed in a broad manner to encompass any record in a database including, but not limited to, a text and or text/image document. The displayed document set may be e.g. the output of a search query that is applied to a search engine (e.g. *AltaVista*TM), or an entire document collection indexed by the system, or any
5 other document set that is provided as an input for displaying to the user in accordance with the invention.

Indexing concepts

10 The documents in the presented document set are characterized by *indexing concepts*, as described above. That is, a typical document is characterized by several indexing concepts that are logically associated thereto. A document is considered *indexed* by the indexing concepts characterizing it.

15 Concept hierarchy

The possible indexing concepts for documents in the system are arranged in a predetermined *hierarchy of indexing concepts* (*hierarchy* in short), as illustrated e.g. in Fig. 5 (31). That is, a *parent* concept (which is an indexing concept by itself) is defined for each indexing concept. For example, in Fig. 5
20 (33) "Countries" is the parent of (34) "Latin America". One or several concepts that are defined as roots of the hierarchy may not have a parent node. For example, in Fig. 3 (32) "All" is the root. Usually, each concept in the hierarchy has only one parent giving the hierarchy the form of a tree data structure (or several trees in case of several roots). The described functionality can
25 accommodate also situations where some nodes have more than one parent. The terms *concept* and *node* are used interchangeably to denote an indexing concept within the hierarchy. Those versed in the art will readily appreciate that the structure of the indexing concept hierarchy is substantially predetermined.

Those versed in the art will readily appreciate that the predetermined
30 structure does not necessarily mean that the indexing categories may not be

subject to modification.

For example, the hierarchy may include an indexing concept
“Companies”, such that some of its specific daughters are not predetermined. The
system may include a mechanism to recognize dynamically that a new name
5 appearing in a document is a company, and define that name as an indexing
concept for the document which is a daughter of the node “Companies”.

By another embodiment, notwithstanding the predetermined nature of the
hierarchy, the system includes a filtering mechanism which in response to
filtering criterion decides whether an indexing concept is displayed, or not, in the
10 hierarchy. For example, the filtering criterion may filter out concepts associated
with a small number of documents, below a certain threshold, or concepts that are
associated only with documents whose score for a search query, whose results list
constitutes the document set to be displayed, is low.

15 *The Concept Hierarchical Display*

In an embodiment of the invention, a system displays the concept
hierarchy (in a hierarchical display) by any visualization mechanism that is
suitable for displaying a hierarchical structure. The most typical display form for
20 a hierarchy is a tree display, as in Fig. 5 (37), in which each node of the tree
corresponds to one concept in the. Clicking on a node (or on a special sign, such
as ‘+’, that is attached to the node) leads to displaying or hiding its daughters.
Other hierarchical display mechanisms may show one level of siblings in the
hierarchy at a time, by showing a list of elements, each represented by some
25 symbol or icon, where clicking on an element leads to displaying its siblings,
while some other option enables getting back (up) in the hierarchy (for example,
the “My Computer” icon in the Windows-98/NT system available from Microsoft
Inc, USA). For the purpose of the invention, any hierarchical display mechanism
can be used to display the hierarchy of indexing concept, where user interaction
30 with the display mechanism controls the display of different portions of the

hierarchy. Another non-limiting example of hierarchical display is a chart, e.g. a pie chart.

Hierarchical Document Set Display

5

This subsection defines a hierarchical display of a presented document set (containing documents indexed by indexing concepts). The hierarchical display serves as a "table of contents" for the document set, which facilitates navigating and browsing of document sets. The scheme of a hierarchical document set display is available in previous systems, but the invention includes some specific enhancements to this scheme, as noted below.

The *hierarchical document set display* is based on the concept hierarchical display, and can be realized by any mechanism for displaying hierarchies, just like the concept hierarchical display discussed above. For example, in Fig. 5 (37) is a hierarchical document set display in tree form. In addition to the predetermined hierarchy of concepts (as explained above) a set of documents, which is a subset of the currently presented document set, is associated with each concept (node) in the hierarchy. In Fig. 5, a set of documents is associated with the node (39) "Countries". The associated document set for a concept in the hierarchy (the document set of the node) contains all documents that are indexed by that concept. In certain embodiments of the invention, the associated document set for a concept is defined to include all documents associated with by any of its decedents in the hierarchy. For example, the document set of the concept "Countries" includes all documents indexed by any country or geographical region, assuming that these concepts are all descendents of the concept "Countries" in the hierarchical display.

It is simple to compute the document set that is associated with a given node in the hierarchical display. As a non-limiting example, such computation may scan all documents in the displayed document set and check for each of them if it is associated with the given concept.

30

A hierarchical document display thus includes a display of the concept hierarchy (as described above), augmented with some information at each concept node about the document set associated with that node. The information about the associated document set may include, by one embodiment, one or more of the following items:

1. The number of documents in the associated set. In Fig. 5 (40) there are 13 documents in the set associated with "Latin America".
2. The percentage (proportion) of documents associated with the concept relative to the number of documents associated with its parent in the hierarchy. In Fig. 3 (41) 7% of the documents in the set associated with "Latin America" relate to "Argentina" (note that a 0% number represents a small positive percentage that was rounded to 0).
3. Some key information about prominent topics described within documents of the document set, such as most frequent or prominent key terms within the documents of the set, and/or the list of all or some of the indexing concepts for the documents.

It should be noted that the nature and form of presenting the specified types of information (by this particular example number of documents, percentage and prominent topics) is only an example and accordingly other types of information may be presented in addition or instead the specified items. Likewise, and as will be explained in greater detail below the concepts and their associated information is not limited to a specific form of graphical and or textual representation.

Reverting now to the specified types, these or other types of information may be presented either textually or graphically. In Fig. 5 (37) is a tree display of the hierarchy with associated information about the document set of each node, containing number of documents and percentage relative to the parent node document set. In particular, since the hierarchical display of indexing concept

may include numerical data, such as numbers and proportions, mechanisms for displaying quantitative information may be used for the display. For example, a pie (or bar) chart can be used to display several sibling nodes (daughters of a common parent). In Fig. 6 (44) is a pie representing the daughter nodes of "Countries". Each pie slice corresponds to one concept and its size indicates the proportion of its associated document set relative to the parent node document set. The quantitative graphical display mechanism may be interactive, in a similar manner to interactive tree presentation of the concept hierarchy. For example, double clicking on a pie slice may lead to displaying the pie of the daughters of the selected node. For example, double clicking on the slice in Fig. 4 (45), corresponding to "Latin America", leads to the display in Fig. 7 (47), a pie presenting the daughters of "Latin America".

The displayed daughters of a node may be sorted alphabetically, or by some characterizing quantitative information, in particular by the size of the associated document set for each daughter.

In accordance with the invention, different display mechanisms are provided. According to the invention, several different display mechanisms may be used interchangeably within a system for the hierarchical document set display, letting the user switch from one to another while maintaining the position within the hierarchy. For example, a system may combine both a pie chart display and a tree display. When viewing the tree display with a certain node selected, and switching to the pie chart display, the system will present the pie that corresponds to the daughters of the selected node.

The graphical display may present further information about the documents in the associated document set, such as their titles, meta-data elements, document summaries or the full text of the document. For example, Fig. 5 (42) is the list of titles for the documents associated with the node "Latin America". Fig. 10 (48) is a summary of a selected document in the document list. A display of the full text of a document is presented in Fig. 8 (52). (54) is a list of indexing concepts for the document.

- 20 -

Optionally, concepts in the hierarchical display to which no documents are attached may be omitted from the display. For example, in Fig. 5 no documents are associated with the indexing concept (36) "Bahamas" in the concept hierarchy, thus in the hierarchical indexing concept display, this concept does not
5 appear as a daughter of (40) "Latin America".

In a more generalized embodiment, the concepts in the hierarchical display are being subject to filtering criterion in order to determine whether or not they will be displayed in said hierarchy. A typical, yet not exclusive, example of filtering criterion concern which folders in deeper levels of the hierarchy tree will
10 be displayed. The necessity of this criterion stems from the fact that the display area allocated to the hierarchy in the display screen may not be sufficient to accommodate the entire hierarchy, and accordingly only portion thereof is displayed, e.g. few levels, and only in response to user selection further levels are displayed (instead of the previously higher levels). For example: if the top level
15 and only some of its daughters are shown, with say a "... " symbol indicating that there are more daughters, that can be displayed if the user explicitly opens the parent node— (as. say in *AltaVista*TM). More advanced filtering criterion may rank folders (standing by this embodiment for nodes) to be presented according to, say the number of documents in it and the quality of their match to the current
20 "query" (query means the entire sequence of operations that led to the display of the current results). Thus, folders having high rank may be displayed in the limited display zone instead of other folders having lower rank, notwithstanding the fact that the higher ranked folders reside in a lower level in the hierarchy as compared to the lower ranked folders. Obviously, the user can display the rest of
25 the folders (which are currently not displayed due to their low rank) by, say, clicking the specified "... " symbol.

According to an embodiment of the invention, an "Others" node is added to each list of siblings having a common parent. By this embodiment, the documents associated with the "Others" node are those associated with the parent
30 node but not with any of its daughters in the concept hierarchy. For example, an

- 21 -

"Others" node that is a daughter of the node "Europe" will be associated with all documents indexed by "Europe" but not by any particular European country.

The hierarchical indexing concept display may be restricted to a particular sub-part of the hierarchy, determined by some mechanism, rather than presenting the full hierarchy. For example, it is possible to present the hierarchical indexing concept display using only the "Countries" sub-tree of the hierarchy. This non-limiting modification also falls in the definition of predetermined hierarchical indexing concept display.

10 *Dynamic Hierarchical Document Indexing Concept Set Display*

The hierarchical indexing concept set display serves as a "table of contents" for the document set and can be used as a method for displaying document sets to the user. However, the hierarchical indexing concept set display is limited because it has a static structure, which is equivalent to the structure of the concept hierarchy. For example, when presenting a large document set by the hierarchical indexing concept display, one of the leaves of the tree may be the country "France", as in Fig. 11 (55), containing 45 documents. No further organization is given for these 45 documents, since "France" is a leaf in the concept hierarchy. This section defines a novel mechanism provided by the invention for presenting dynamic "tables of contents" displays for document sets, enabling the user to dynamically modify and refine the document display whilst maintaining the predetermined hierarchical indexing concept display. This mechanism is called the *dynamic hierarchical document set display (dynamic display)*. The dynamic display is by itself hierarchical utilizing the specified predetermined hierarchy of categories, and thus provides all the functionality of the hierarchical document set display, as described above..

30 In accordance with one embodiment, at the initial stage of the dynamic

display, a document set is presented in some manner, possibly by the (static) hierarchical indexing concept display. The dynamic display is created by a series of “*organize by*” operations, each specified by two definitions:

- 5 1. Defining a document subset (or set), to be organized (constituting
 “organized” document subset) by the “organize by” operation. For a
 hierarchical presentation, selecting the document set may, preferably,
 correspond to selecting a node in the hierarchical document set
 display. For example, selecting the node “France” in Fig. 11 (55)
10 defines the document set associated with this node as the subset to be
 organized. This subset is termed the organized document subset.
 When the selected subset corresponds to a node in the display, that
 node is termed the organized node. The selection of the “organized”
 document subset is performed on the basis of information displayed
15 in the hierarchy, e.g. defining an indexing concept in the hierarchy as
 an organized by concept and rendering the documents associated
 therewith as the specified “organized” document subset.
2. Defining a node of the concept hierarchy to serve as the root of
 the sub-tree by which the document subset will be organized. This
20 node (or corresponding sub-tree) is termed the organizing node
 (sub-tree). For example, the node “Companies” may be selected as
 an organizing node (57 in Fig. 9), to organize the document subset
 associated with the node “France”.

25 The effect of applying the “organize by” operation is to provide an
 “organizing” hierarchical indexing concept display (as defined above) for the
 organized document subset, which is restricted to the sub-hierarchy under the
 organizing node. In the above example, the documents associated with the node
 “France” will be displayed in a hierarchical indexing concept display that is

- 23 -

restricted to the sub-tree of the concept hierarchy rooted by the node "Companies" (having all companies as daughters). This display appears in Fig. 9 (60), where the "Companies" node (60) is the root of the hierarchical display for the "France" document set, and (58) are the daughters of (60). This would have the effect of presenting which companies appear as indexing concepts in documents that are also indexed by "France", along with quantitative information about the documents indexed by each company. For example, there are 27 documents indexed by both "France" and "Boeing". The indexing concept Boeing (61) signifies, due to its position in the hierarchy, the path from the root to wit: All->countries->West Europe->France->Companies-> Boeing. Put differently, indexing concept (61) is associated with the documents indexed by both "France" (a country in west Europe) and "Boeing (company). The pertinent information that is associated with this concept is 27 (No. of documents) and 60% (standing for 27 documents out of the 45 associated with indexing concept (60) – Boeing. Accordingly, any concept in the indexing concept hierarchy display is associated with respective sub set of documents from among the organized document subset. Obviously a document may be associated with more than one concept of the organizing hierarchical display. A "respective" subset of documents encompasses also the special situation in which a concept is associated with no documents.

The "organize by" operation may be interpreted as a recursive application of the hierarchical indexing concept display, as its effect is to provide a new hierarchical display for a node within a previously displayed hierarchy. However, the hierarchical display is maintained predetermined considering that in the modified presentation, substantially, the same concepts are employed, which makes it easier for the user to follow "well known" and familiar concepts, even after applying the "organizing" operation.

As a special case, the organizing node can be the root of the concept hierarchy, in which case the organized document subset will be displayed by a hierarchical indexing concept set display that corresponds to the entire concept

hierarchy. A system may apply only this special case (always organizing by the full hierarchy considering the root as the organizing node), in which case it is necessary to define only the organized node in order to apply an "organize by" operation. Furthermore, a system may implement the hierarchical document display such that at each point of time the user view is focused only on one node of the tree. In this case, applying the "organize by" operation implies implicitly that the organized node is the currently displayed node, saving the need of an explicit definition of the organized node. If desired, by a specific embodiment, the default definition of organizing concept as the root node and the organized by concept as the currently displayed node may be realized by a single user operation say, for example, clicking on a predetermined icon.

As a particular (but not the only) mechanism of operation, in the case where the organized document subset corresponds to a specific selected node in a hierarchical display, the hierarchical display of the organized subset is displayed as a new, dynamically created, daughter (or daughters) of the selected organized node. In the example above, the node "Companies" in Fig. 7 (60) is added dynamically as a new daughter node of (59) the node "France", modifying the hierarchical display that was presented to the user just before applying the "organize by" operation. Several variations of the method may be implemented, in which a new daughter node either replaces or is added as a sibling to the previously existing daughters of the organized node. Notwithstanding the modification, the predetermined hierarchy of concepts is maintained in the sense that the category "company" is already known to the user (see e.g. Fig. 3) before applying the specified "organize by" operation.

Once a modified hierarchical display has been created by applying an "organize by" operation, as described above, any part of the new display may be subject to further "organize by" operations. In particular, a node that was added to the hierarchy in a previous "organize by" operation may be selected as the organized subset in a later operation. Subsequent "organize by" operations on the modified dynamic display may be applied as requested by the user. In Fig 13 (69)

- 25 -

the node "Boeing" which has been created by a previous "organize by" operation (as in Fig. 9) is later selected as an organized node, where the organizing node is (65) "Activities". Thus, in this example, a node "Activities" (70) is dynamically added to the display, and its daughters (64) (signifying documents indexed by both "France" and "Boeing" and by some activity) are associated, each, with information that pertains to these documents. For example, there are 19 documents indexed by "France" (67) "Boeing" (69) and "Agreement" (71). The specified organized by operation may be applied recursively (repeated) as many time as required each time in respect of new selected "organized by" and "organizing" concepts.

The basic form of the "organize by" operation may consist selecting one node in a hierarchical display as the organized node, and one node in the concept hierarchy as the organizing node. The following paragraph describes extensions to the basic form.

Multiple selection for simultaneous operation

Multiple selection of organizing nodes within a single "organize by" operation has the effect by one embodiment of adding all the selected nodes as daughters of the organized node. For example, the organizing node "France" may be organized by, "Companies" and "Activities", which means that all the documents associated with the indexing concept France will be organized by the indexing concept "Companies" and separately by the indexing concept "Activities". If desired, the nodes "Companies" and "Activities" are added as daughters to "France".

Multiple selection of organized nodes has the effect of applying the "organize by" operation simultaneously to all selected nodes. For example, applying an "organize by" operation with the same organizing node to both nodes "France" and "Spain". The net effect of selecting more than one organized nodes is that each node is associated with its respective organized by subset of documents and

then some operator or operators is (are) applied to the specified subsets so as to constitute resulting organized subset of documents that is then subject to the organizing operation. In the latter example there is a first subset of documents associated with *France*, a second subset of documents associated with *Spain*. By this particular example the operator that is applied to the subsets is *OR* giving rise to a document subset that includes documents that pertain only to Spain, only to France or to both. This resulting subset of documents is then being subject to the organizing operation by one or more organizing concepts.

In accordance with an embodiment of the invention, the set of documents may be obtained by applying a search query to say conventional search engine that operates similarly to as AltaVistaTM and display the resulting set along with the hierarchical display of the invention.

Thus, for example, Figs. 14 to 21 showing a succession of screen results by applying the method in accordance with one embodiment of the invention.

Fig. 14 illustrates a predetermined indexing concept hierarchy (140) that includes 11,000 documents (142) that constitute the document set and are broken down by the hierarchy concepts.

Applying a query (e.g. *paggers* 143) results in 318 documents (see 151 in Fig. 15) that are broken down by the concept hierarchy. The list of documents is displayed (152), and, by this example, the first four documents are shown in the first page. The query itself ("paggers") is automatically assigned to categories in the hierarchy as if it were a document. The resulting category is illustrated in the "Related category" field (153), to wit: Telecom All > Applications > Messaging > Paging. All the categories, except from "Paging" are shown in the hierarchical presentation (151, 154, and 155). *Paging* is a sub category of *Applications* and can be shown if the *Browse* section of the screen is enlarged, or if the user decides to show it by, say, clicking a specified symbol (as described above).

Clicking the *Paging* will render the latter *organized by* category and the *All* (i.e. the root) organizing category. The net effect is that the 122 documents that are associated with *Paging* are now broken down by the entire hierarchical

tree, as shown in Fig. 16. The "results for" field shows that the display corresponds to the query "paging" (which by this example matches one of the categories). The four documents shown in the search section are the first 4 out of 122 documents that meet the search.

5 Fig. 17 is the same as Fig. 16 except that now the documents that are associated with sub-category Telecom Service Companies (171) are shown. This may be achieved by simply clicking the relevant category in the hierarchy (by this particular example *Telecom Service Companies* – not shown in the hierarchy Fig. 17) and the documents associated therewith are shown. The documents that are
10 shown obviously relate to "paging" and telecom service companies.

Fig. 18 illustrates yet another degree of detail wherein only documents that pertain to *SkyTel* 181 (which forms sub-category of the specified *Telecom Service Companies* – not shown in the hierarchy of Fig. 18) are shown.

Now, Skytel constitutes the organized by concepts and the documents
15 associated therewith constitute the organized document subset. Next, clicking the *Zoom In* symbol (182) will render the *Telecom All* root category (183) the organizing category and the resulting hierarchical display is depicted in Fig. 19.

There are 12 documents (191) broken down by the predetermined categories. Thus, for example, 8 documents are associated with the category
20 Business (192). Categories that have no documents associated therewith are not shown. Incidentally, the information that pertains to the sub documents associated with each category is simply the number of documents (12 and 8 in the latter example). The 12 documents concern both *Skytel* and *paging*. Four out of the 12 pertinent documents are shown in the Search section of the
25 screen (193).

Considering now that only the documents from among the specified 12 documents that concern product companies (the "products" node) are of interest, the user simply clicks the *products* category (200) in Fig. 20 and the 8 relevant documents are shown at the search section of the screen (201)

30 If, from among the specified 8 documents only those that concern

Motorola are of interest the user simply clicks the *Motorola* category (210) in Fig. 21 and in response thereto the pertinent 3 documents are shown.

Selecting text terms and segments for focused reading

5

In addition to the dynamic display, which provides a "table of contents" style display for document sets, the invention provides in accordance with another aspect thereof, new mechanisms for presenting parts of or all of the text of a document in a dynamic and effective manner. These mechanisms direct the
10 attention of the user to relevant parts of the document and enables quick focusing on these parts. For example, these relevant parts might be text segments that contain relevant information for the user or can help deciding about the relevance of the document. The decision of which parts of the document should be in focus is dynamic, and may be changed according to user guidance or to the context in
15 which the document is being displayed.

There are two typical ways in an embodiment of the invention for focusing the user attention on particular parts or pieces of information in the document. The first is by highlighting the parts of the text that should be in focus (based on important triggering terms) , and the second is by creating a summary for the
20 document that contains the parts in focus (based on important triggering terms).

According to the invention, the parts of the document which should be highlighted or be included in a summary are determined according to a set of (one or more) indexing concepts, among the indexing concepts of the document, that are considered to be in focus at a certain stage of user interaction with the system.
25 These indexing concepts are called *focus indexing concepts*.

According to the invention, the highlighting and summarization for a given focus indexing concept is determined by *the important triggering terms* for that concept. The triggering terms for a concept are the occurrences in the document of all terms which entail the attachment (or classification) of the
30 concept to the document. Highlighting and an extracted summary will include the

important triggering terms for the concept, or short segments of text that are considered to be important. The degree of importance of terms and segments may be quantified by some scoring mechanism, where the degree of importance of the terms in a segment is factor in determining the degree of the segment importance.

- 5 The invention provides dynamic methods for determining (quantifying) which triggering terms and segments are important in a given context of the user interaction with the system that displays the documents.

It should be noted that in the context of obtaining a summary according to one embodiment of the invention, the quantifying step assigns the same degree of
10 importance to all triggering terms. The latter option does not apply to the aspect which concerns emphasizing important triggering terms. Put differently, insofar as emphasizing important triggering terms, not all the triggering terms are ranked with the same degree of importance.

The important triggering terms and segments are presented to the user,
15 either in a form of an extracted summary, which contains the important terms and/or segments, or by highlighting the important terms within the display of the full document, or by some combination of the two methods. When using the term *important*, one refers to the case where the degree of importance of triggering terms and segments can be quantified and the display is restricted those with the
20 highest importance. The amount of terms or segments to be included in the display is determined by some mechanism, such as a threshold on the degree of importance or on the number of items to be included. This ranking mechanism by degree of importance is necessary when there are many important terms or segments and it is desired to limit their display in order to achieve optimal focus
25 of attention by the user. Fig. 10 (48) displays a summary of a document, in which the important terms are highlighted. (The important terms were determined relative to the highlighted indexing concepts "Latin America" (50) and "Lockheed Martin" (51) which are in the focus of interest to the user, as explained below). The summary includes segments of the text that contain the
30 important terms. Fig. 8 (52) presents a full display of a document text, in which

important terms (relative to the indexing concept (54), see below) are highlighted. While the general scheme of making some form of highlighting triggering terms in a document for display is available in previous systems, the invention, by this aspect, concerns selecting important terms, described below.

5

Selecting the important triggering terms within a text classification system that quantifies the importance of triggering terms

One non-limiting method in the context of the invention refers to selecting
10 the important triggering terms in a document with respect to an indexing concept that is determined to be in *focus* (of interest) at a certain stage of the user interaction with the system. For example, in Fig. 8 the indexing concept "Product specifications/capabilities" (54) is selected to be in focus. This part of the invention refers to the case where the indexing concept was assigned to the
15 document by some text classification method, as described above. Such a method classifies the document to a certain indexing concept based on words, terms or their combinations that appear in the document. It is assumed that it is possible to trace within the classification system which words or terms in the document entailed the classification to the given indexing concept. Optionally, it is possible
20 to quantify within the system the relative contribution of each term to the classification of the document to the indexing concept. In certain embodiments, a trainable text classification method in which the terms and the degree to which they entail classification to the indexing concept are learned from training documents, for which it is previously known whether they belong to the indexing
25 concept or not.

As non-limiting examples for possibilities for determining triggering terms, consider the following trainable text classification methods.

- D. Lewis, 1992, An evaluation of phrasal and clustered representations on a text categorization problem, in Proc. of the 15th Int. ACM-SIGIR
30 Conference on Information Retrieval, pages 37—50. This method

- 31 -

applies a Bayesian learning scheme for text classification. For a given category, the method computes (during the training phase) certain weights for terms (words or phrases) in the text, with respect to the category. The score of the category for a particular document is
5 computed as a function (usually some sort of a normalized sum) of the weights of the terms that appear in the document. When computing the category score for a document, it is possible to trace the relative contribution of each term in the document to the accumulative score. Thus, triggering terms in this method will be those terms that provided
10 the highest contribution to the accumulative score of the document.

- E. Wiener and J. Pedersen and A. Weigend, 1995, A neural network approach to topic spotting, in Symposium on Document Analysis and Information Retrieval, pages 317—332.
- W.W. Cohen, Text categorization and relational learning, in Machine
15 Learning Journal, 1995, pages 124—132. This method learns classification rules for each category, that consist of words or combination of words. Each “firing” of a rule, that is, the occurrence of the word or word combination of the rule in the document, entails the classification of the document to the category. Thus, in this method, the words and word
20 combinations in the rules that matched in the document will be considered as triggering terms in the document.

According to the invention, the important triggering terms, to be included in a summary or to be highlighted, are those term occurrences that significantly
25 contributed to the classification of the document to the focus indexing concept. In Fig. 8 the triggering terms for the indexing concept “Product Specification/Capabilities” (54) are highlighted within the text (52) of the document. Furthermore, when the relative contribution of triggering terms to

classification can be determined (traced) then their degree of importance would be proportional to this degree of relative contribution to classification.

It should be noted that the method described above for selecting the important triggering terms for an indexing concept in focus could be combined
5 with simpler methods for identifying the triggering terms for an indexing concept (such methods are not part of the invention). For example, when the indexing concept is identical to a term or name that appears explicitly in the document text then the important term is simply the occurrence of the indexing concept in the text. (E.g. when the indexing concept is "France" and the important terms are
10 simply the explicit occurrences of the term "France" in the text). Another example is a topical indexing concept that is identified in the text by a manually defined query. In this case the triggering terms are simply all the terms that appear in the query (similar to document search systems that highlight matching query terms in the retrieved documents).

15 Those versed in the art will readily appreciate that the invention is not bound to the specified specific techniques for determining important triggering terms.

Multiple focus indexing concepts

Another method within the invention refers to selecting important terms
5 and segments for display by selecting dynamically several focus indexing
concepts. One way of selecting the focus indexing concepts is by letting the user
select them interactively from the list of all indexing concepts of the document. In
Fig. 10 the user have selected (50) "Latin America" and (51) "Lockheed Martin"
as focus indexing concepts. Consequently, the selected important terms, which
10 are highlighted in the document text (48), are the triggering terms for both (50)
and (51). Other mechanisms for selecting the set of focus indexing concepts may
be applied as well, such as the method described next. According to the
invention, the important triggering terms and segments are selected from the
important triggering terms and segments of each one of the focus indexing
15 concepts; applying some procedure that combines them and reevaluates their
degree of importance with respect to the complete set of focus indexing concepts.
For example, the degree of importance of a triggering term or segment with
respect to the complete set of focus indexing concepts may be defined (referred to
also as quantified) by its maximal (or minimal) degree of importance for any of
20 the individual indexing concepts (applying a disjunctive (or conjunctive)
reasoning criterion), or by computing some averaging function of the individual
importance degrees. According to the invention, the display of important terms or
segments for the complete set of focus indexing concepts may distinguish
between terms that were selected originally for the different indexing concepts
25 that compose the set. For example, a different color is attributed to each indexing
concept, and the important terms related to this concept are highlighted by the
corresponding color. In Fig. 10 the indexing concept "LATIN AMERICA" (50)
is highlighted with a blue background and "LOCKHEED MARTIN" (51) is
highlighted with a pink background (blue appear darker than pink in the black
30 and white printing). Accordingly, the triggering terms for both concepts ("Brazil"

- 34 -

and "Amazon" for "LATIN AMERICA" and "Lockheed Martin" for the indexing concept "LOCKHEED MARTIN") are highlighted in the corresponding colors in the document text (48).

5 Default focus indexing concepts

Another method within the invention refers to the selection of *default focus indexing concepts*, to be used automatically as the focus indexing concepts when the document is presented to the user. According to the invention, the default focus indexing concepts are selected according to the selection conditions that were applied in the process that led to the display of the document. In particular, when the document is displayed as a result of a search query that contains indexing concepts then the indexing concepts contained in the query become the default focus indexing concepts. A particular setting for this method occurs when the document is selected for display within the hierarchical document set display or within the dynamic hierarchical document set display. In Fig. 12 a document was selected for display from the node (document subset) (61) "ARGENTINA". Accordingly, the *default focus indexing concept* is (62) "ARGENTINA" and the triggering term "Argentina" (63) is highlighted within the document text.

In this setting of a hierarchical display a document is selected for display from the document set that is associated with a certain node in the hierarchy. The documents in this set satisfy a logical condition that is equivalent to a search query which is a conjunction (logical AND) of all indexing concepts in the path from the root of the displayed hierarchy to the selected node. Thus, according to the invention, the default focus indexing concepts are the concepts along this path. Recall that parts of this path may correspond to paths within the concept hierarchy and parts of the path might be created dynamically within the dynamic hierarchical document set display. For example, in Fig. 13 the documents associated with the node "Agreement" (71) satisfy a logical AND condition for

- 35 -

all indexing concepts on the displayed path from the root of the tree to this node. Optionally, for a pair of concepts x and y in the set of default focus indexing concepts, such that x is an ancestor of y in the concept hierarchy, it is possible to exclude x from the set of default focus indexing concepts. In the example of Fig. 11, it is possible to exclude "West Europe" from the set of default focus indexing concepts since it is likely that the focus of interest for the user is concerned in particular with "France", which is a daughter of "West Europe" in the concept hierarchy.

In some systems, the method of viewing document sets that are attached to concept nodes in a (possibly dynamic) hierarchical document set display may be combined with the use of explicit search queries issued by the user. In this case, if the document set attached to a concept node is restricted by an additional condition supplied in an explicit search query, then the default focus indexing concepts will be a combination of the concepts of the path, as described above, and the concepts that are included in the query.

Alphabetical characters and Roman symbols are designated in the description below for convenience only and do not necessarily imply a particular order of the method steps.

The present invention has been described with a certain degree of particularity, but those versed in the art will readily appreciate that various alterations and modifications will be carried out without departing from the scope of the following Claims. Thus, by way of example, whereas, typically, the organized document subset is determined by defining one (or more) of the concepts in the hierarchy as "organized by" concept, thereby rendering the subset of documents associated therewith "organized document subset" this is not necessarily always the case. Thus, according to a more generalized embodiment any determination of subset of documents (organized document subset) by utilizing the so displayed hierarchy (i.e. implemented using information derived from the so displayed hierarchy) is embraced by the invention.

30

CLAIMS:

1. A method for dynamically presenting set of documents to users , comprising:

- (a) providing a predetermined hierarchy of indexing concepts;
- 5 (b) providing a set of documents;
- (c) providing hierarchical display of the indexing concepts such that each indexing concept in the hierarchy is associated with a sub-set of documents from among said set of documents;
- 10 (d) applying steps that include the following (i) to (iii) , as many times as required:
 - (i) determining a subset of documents by utilizing the hierarchical display, thereby rendering it organized document subset;
 - 15 (ii) defining at least one indexing concept in the hierarchical display so as to constitute a respective organizing concept; and
 - (iii) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the
20 respective at least one organizing concept; each concept in said at least one organizing hierarchical display is
25 associated with a respective subset of documents, from among said organized document subset.

2. The method of Claim 1, wherein said "organizing" concept being the root concept of said hierarchical display.
3. The method according to claim 9, wherein said "organized by" concept is the
5 concept on which the hierarchical display is focused at a given time.
4. The method according to anyone of the preceding Claims, wherein said steps (d)(i) and (d)(ii) are obtained by activating a single command.
5. The method according to anyone of the preceding Claims, wherein said hierarchical display is in a form of tree.
- 10 6. The method according to anyone of Claims 1 to 4, wherein said hierarchical display is in a form of a chart.
7. The method according to Claim 6, wherein said chart and tree representations are interchangeable whilst maintaining the position in the hierarchical display.
- 15 8. The method according to anyone of the preceding Claims, wherein concepts having no documents associated therewith are not displayed.
9. The method according to anyone of the preceding Claims, wherein said step (d)(i) includes: defining at least one indexing concept in the hierarchical display so as to constitute a respective "organized by" concept; the documents associated
20 with said organized concept constitute organized document subset.
10. The method according to anyone of the preceding Claims, further including "others" concept in at least one position in said hierarchical display.
11. The method according to anyone of the preceding Claims, further comprising the step of:
25 displaying at least one desired document or portion thereof from among the documents.
12. The method according to Claim 11, further comprising the step of: displaying

- 38 -

said document with emphasis on important triggering terms that correspond to default focus indexing concepts.

13. The method according to Claim 12, further comprising the step of: obtaining a summary based on said important triggering terms.

5 14. The method according to Claim 12, wherein said emphasis being highlighting the important triggering terms in a predetermined color.

15. The method according to anyone of the preceding Claims, further comprising the step of applying a filtering criterion in said steps (c) in order to determine the concepts that will be in said hierarchical display.

10 16. The method according to anyone of the preceding Claims, further comprising the step of applying a filtering criterion in said steps (d)(iii) in order to determine the concepts that will be in said organizing hierarchical display.

17. The method according to anyone of the preceding Claims, comprising the following preliminary step of: applying a search query to a search engine and
15 obtaining as a result said set of documents, stipulated in said step (b).

18. The method according to Claim 17, further comprising the step of displaying said set of documents in a displaying format of said search engine.

19. A method for presenting set of documents to users comprising

- 20 (a) providing indexing concepts and a set of documents ;
the set of documents are associated with concepts in
accordance with triggering terms in said documents;
- (b) selecting a document from said set;
- (c) selecting at least one concept associated with said
document;
- 25 (d) quantifying the importance of the triggering terms of
said at least one concept and in response thereto,
determining the important triggering terms; and
- (e) emphasizing the important triggering terms that

- 39 -

correspond to said at least one concept.

20. The method according to Claim 19, wherein said emphases being highlighting the important triggering terms in a color that corresponds to
5 respective indexing concept.

21. The method according to Claim 19, further comprising the step of: obtaining a summary based on said important triggering terms.

22. A method for presenting set of documents to users comprising

- 10 (a) providing indexing concepts and a set of documents ;
the set of documents are associated with concepts in
accordance with triggering terms in said documents;
- (b) selecting a document from said set;
- (c) selecting at least one concept associated with said
document;
- 15 (d) quantifying the importance of the triggering terms
of said at least one concept and in response thereto,
determining the important triggering terms; and
- (e) obtaining a summary based on said important
triggering terms.

20 23. The method according to Claim 22, wherein said quantifying step renders
all the triggering terms, as important triggering terms.

24. A method for dynamically presenting set of documents to users, comprising

- (a) providing a predetermined hierarchy of indexing
concepts;
- 25 (b) providing a set of documents;
- (c) providing hierarchical display of the indexing
concepts; the indexing concepts are associated
with the set of documents

- 40 -

- (d) determining a subset of documents by utilizing the hierarchical display, thereby rendering it organized document subset;
- (e) defining at least one indexing concept in the hierarchical display so as to constitute a respective "organizing" concept; and
- (f) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the respective at least one organizing concept, wherein concepts in said organizing hierarchical display are associated with the organized document subset;
- (g) repeating steps (d) to (f), as many times as required.

25. The method according to Claim 24, wherein said step (d) includes: defining at least one indexing concept in the hierarchy so as to constitute a respective "organized by" concept; the documents associated with said organized concept constitute organized document subset;

26. The method according to Claim 25, comprising the following preliminary step of: applying a search query to a search engine and obtaining as a result said set of documents, stipulated in said step (b).

27. The method according to Claim 26, further comprising the step of displaying said set of documents in a displaying format of said search engine.

28. A system that includes a processor associated with a memory and display for dynamically presenting set of documents to users, comprising:

- (a) the memory is configured to store of a predetermined hierarchy of indexing concepts;
- (b) the memory is configured for store a set of documents

- 41 -

- 5 (c) the processor is configured to provide hierarchical display of the indexing concepts such that each indexing concept in the hierarchy is associated with a sub-set of documents from among said set of documents;
- (d) the processor is configured to apply steps that include the following (i) to (iii) , as many times as required:
- 10 (i) determining a subset of documents by utilizing the hierarchy of display, thereby rendering it organized document subset;
- (ii) defining at least one indexing concept in the hierarchy of display so as to constitute a respective "organizing" concept; and
- 15 (iii) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the respective at least one organizing concept; each concept in said at least one
- 20 organizing hierarchical display is associated with a respective subset of documents, from among said organized document subset.
- 25 29. A system that includes a processor associated with a memory and display for presenting a set of documents to users comprising:
- (a) the memory is configured to store indexing concepts and a set of documents ; the set of documents are associated with concepts in accordance with

- 42 -

- triggering terms in said documents;
- (b) the processor is configured to select a document from said set;
 - (c) the processor is configured to select at least one concept associated with said document;
 - (d) the processor is configured to quantify the importance of the triggering terms of said at least one concept and in response thereto, determining the important triggering terms; and
 - (e) the processor is configured to emphasize in the display the important triggering terms that correspond to said at least one concept.

30. A system that includes a processor associated with a memory and display for presenting set of documents to users comprising:

- (a) the memory is configured to store indexing concepts and a set of documents ; the set of documents are associated with concepts in accordance with triggering terms in said documents;
- (b) the processor is configured to select a document from said set;
- (c) the processor is configured to select at least one concept associated with said document;
- (d) the processor is configured to quantify the importance of the triggering terms of said at least one concept and in response thereto, determining the important triggering terms; and
- (e) the processor is configured to obtain a summary in said display based on said important triggering terms.

31. A system that includes a processor associated with a memory and display for

- 43 -

dynamically presenting set of documents to users, comprising:

- (a) the memory is configured to store a predetermined hierarchy of indexing concepts;
 - 5 (b) the memory is configured to store a set of documents;
 - (c) the processor is configured to provide hierarchical display of the indexing concepts; the indexing concepts are associated with the
10 set of documents;
- the processor is configured to apply the following steps (d) to (f) as many times as required.
- (d) determining a subset of documents by utilizing the hierarchical display, thereby
15 rendering it organized document subset;
 - (e) defining at least one indexing concept in the hierarchical display so as to constitute a respective "organizing" concept; and
 - 20 (f) providing at least one organizing hierarchical display of indexing concepts, wherein the root of said at least one organizing hierarchical display being the respective at least one organizing concept, wherein concepts in said organizing hierarchical display are associated with the organized document
25 subset.

1/21

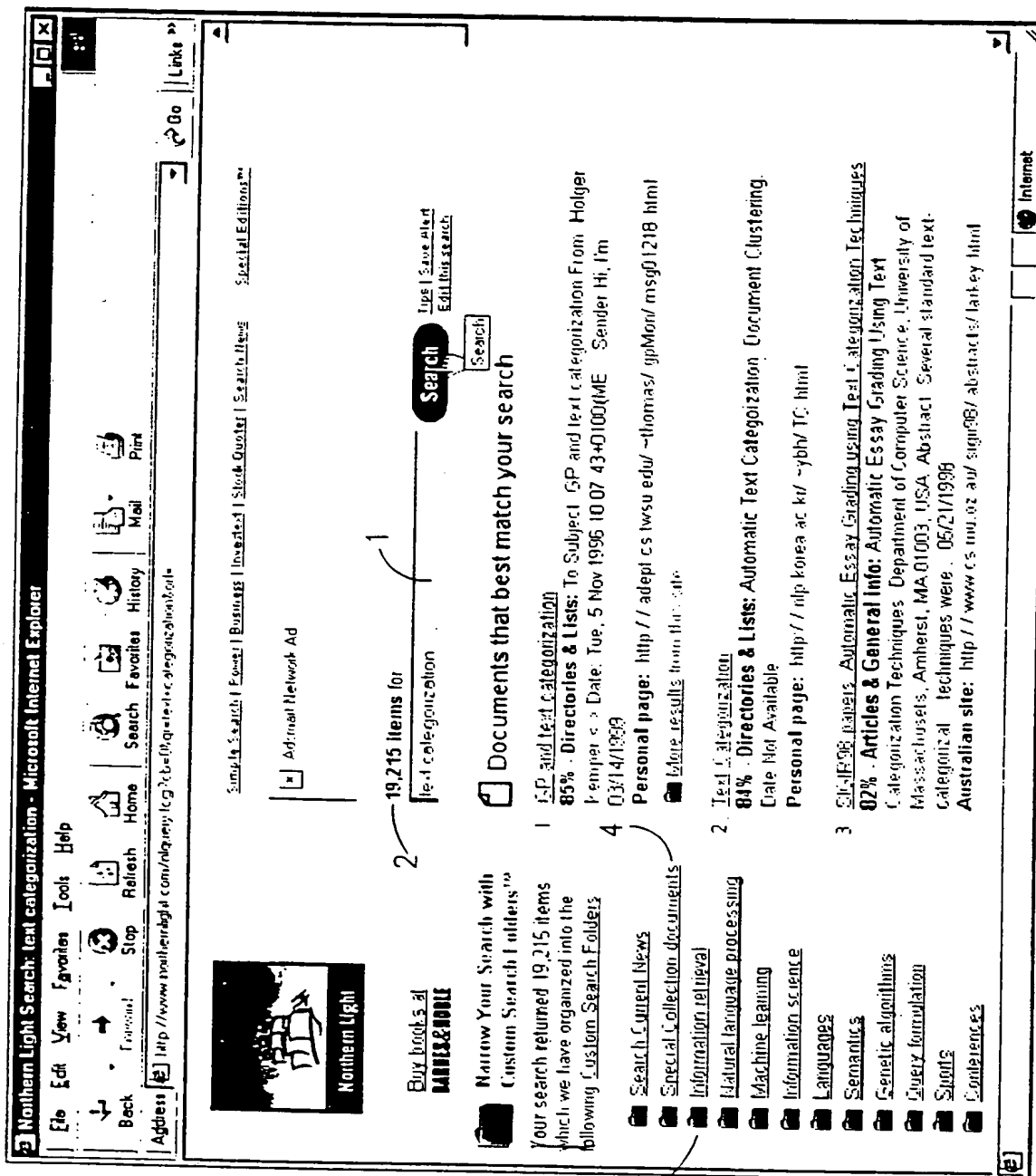


FIG. 1

2/21

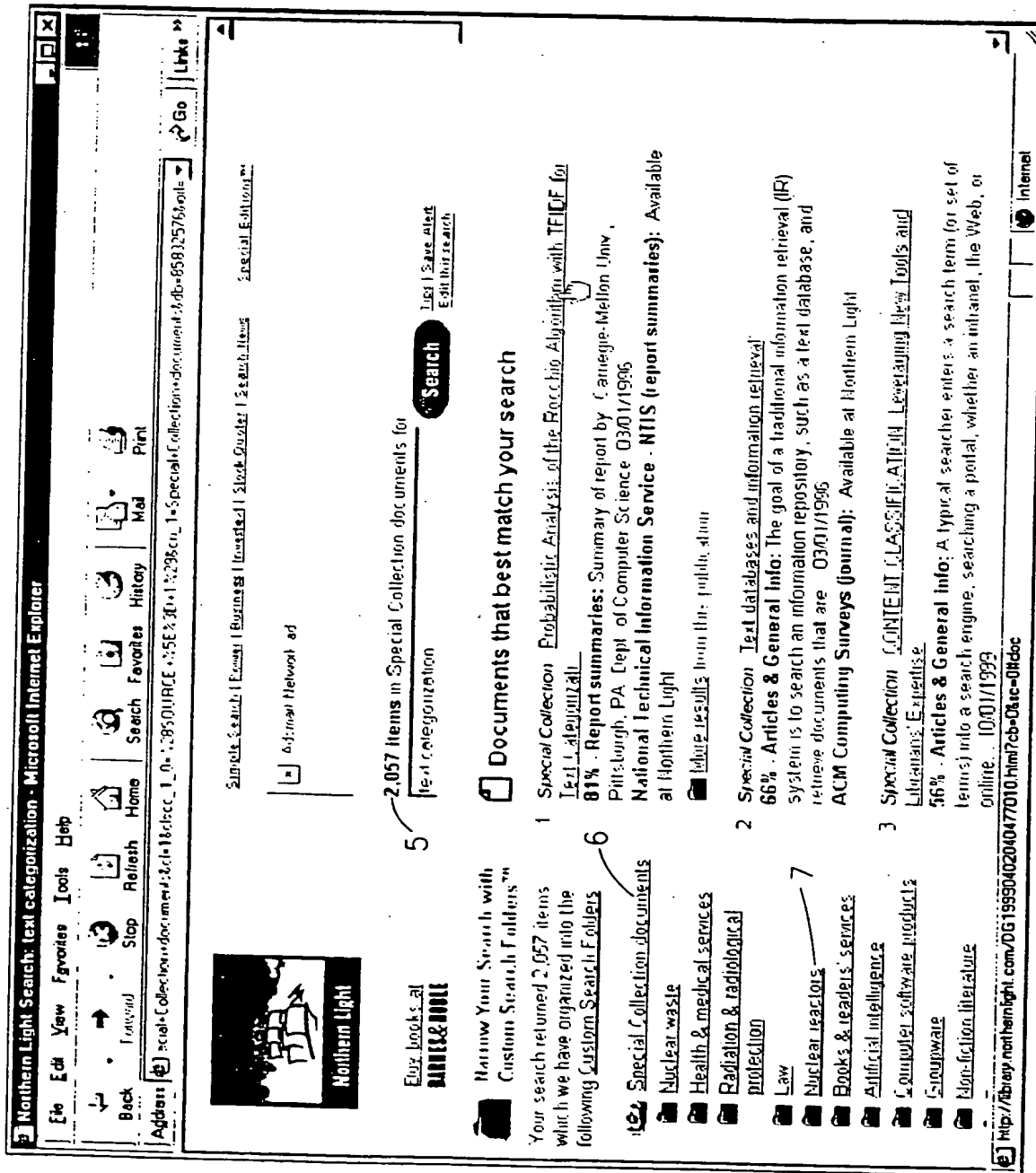


FIG.2

3/21

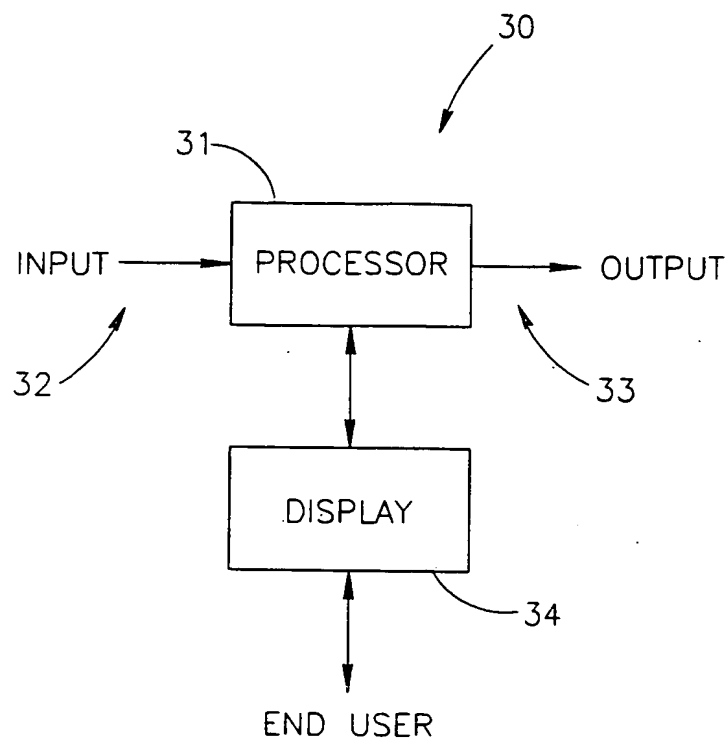


FIG.3

4/21

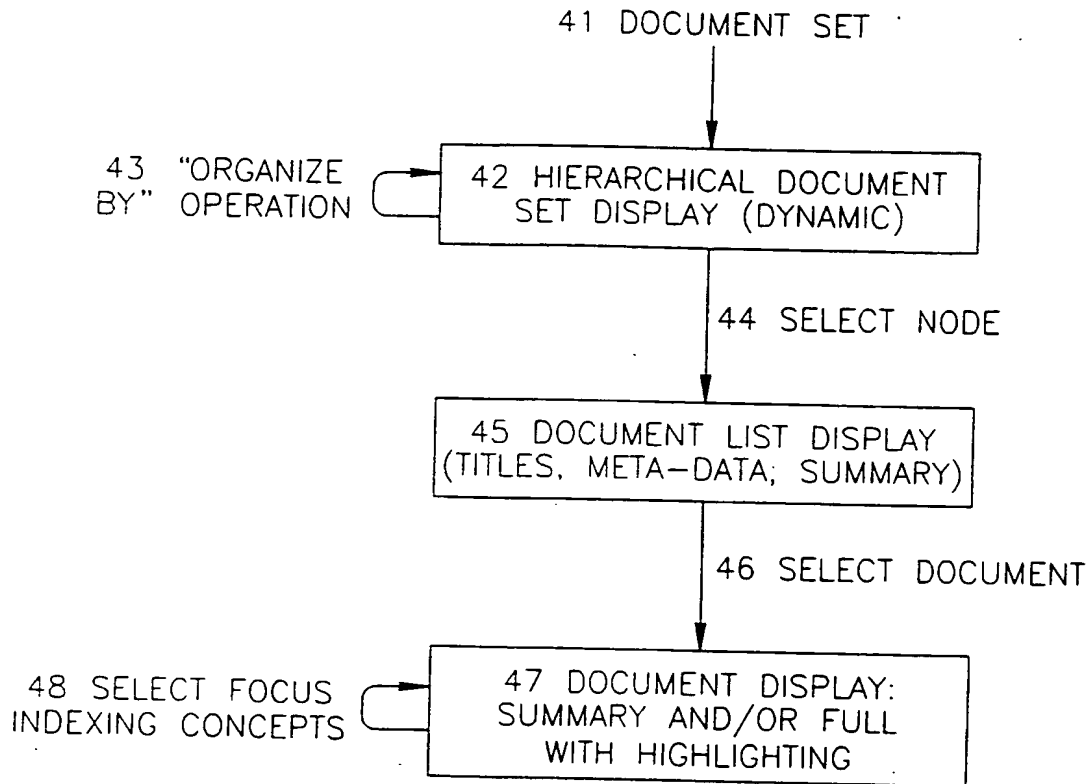


FIG. 4

5/21

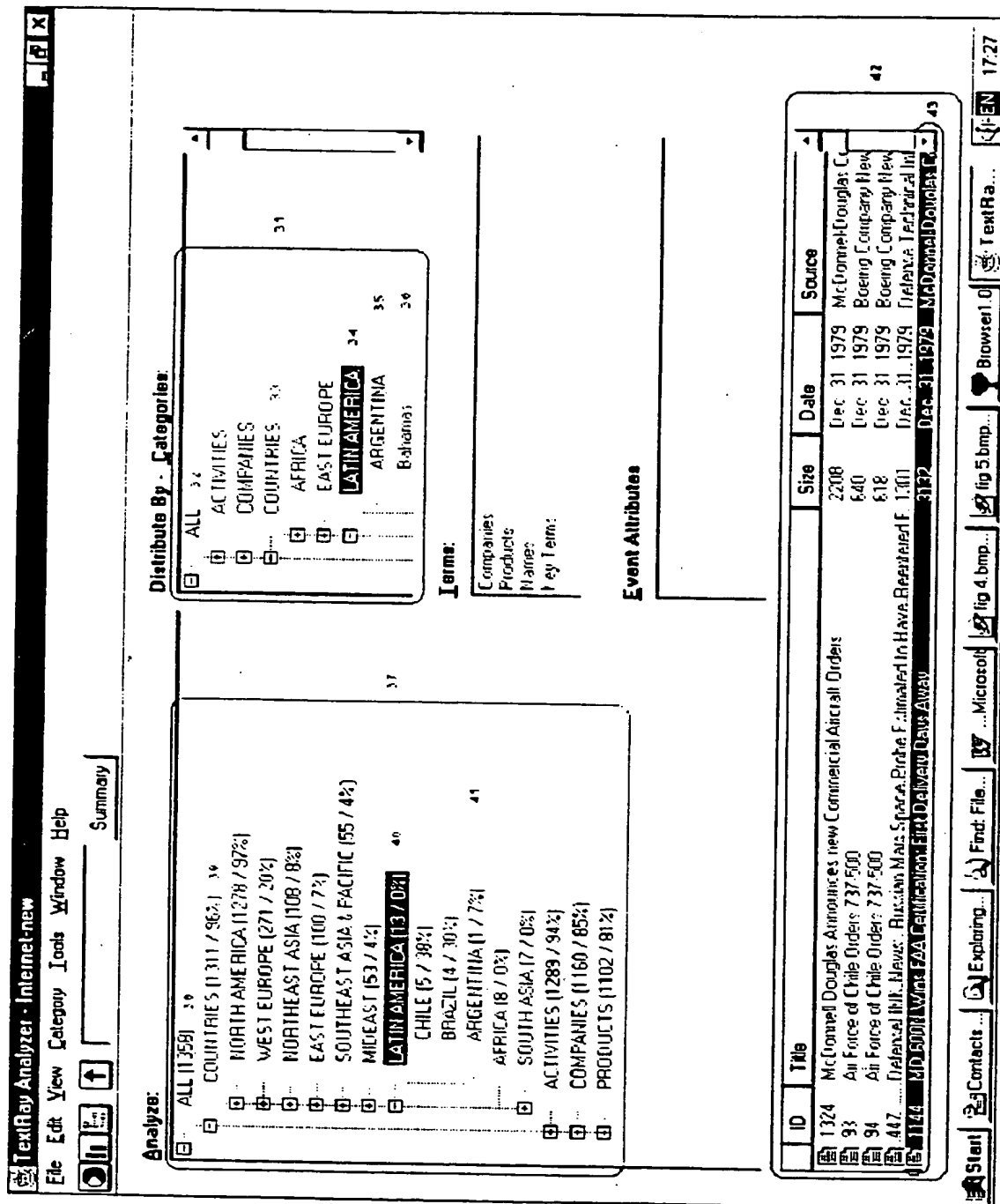


FIG.5

6/21

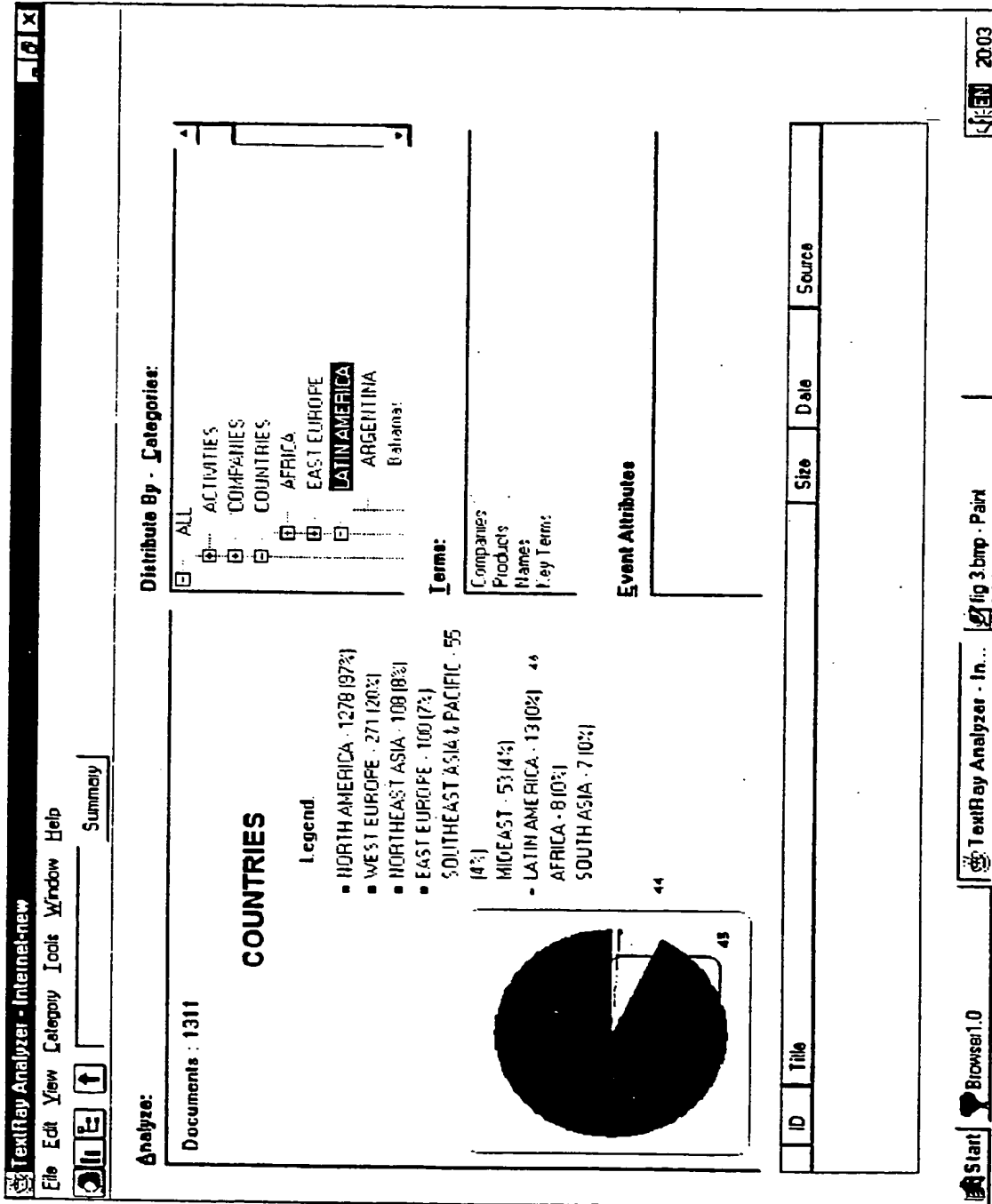


FIG. 6

7/21

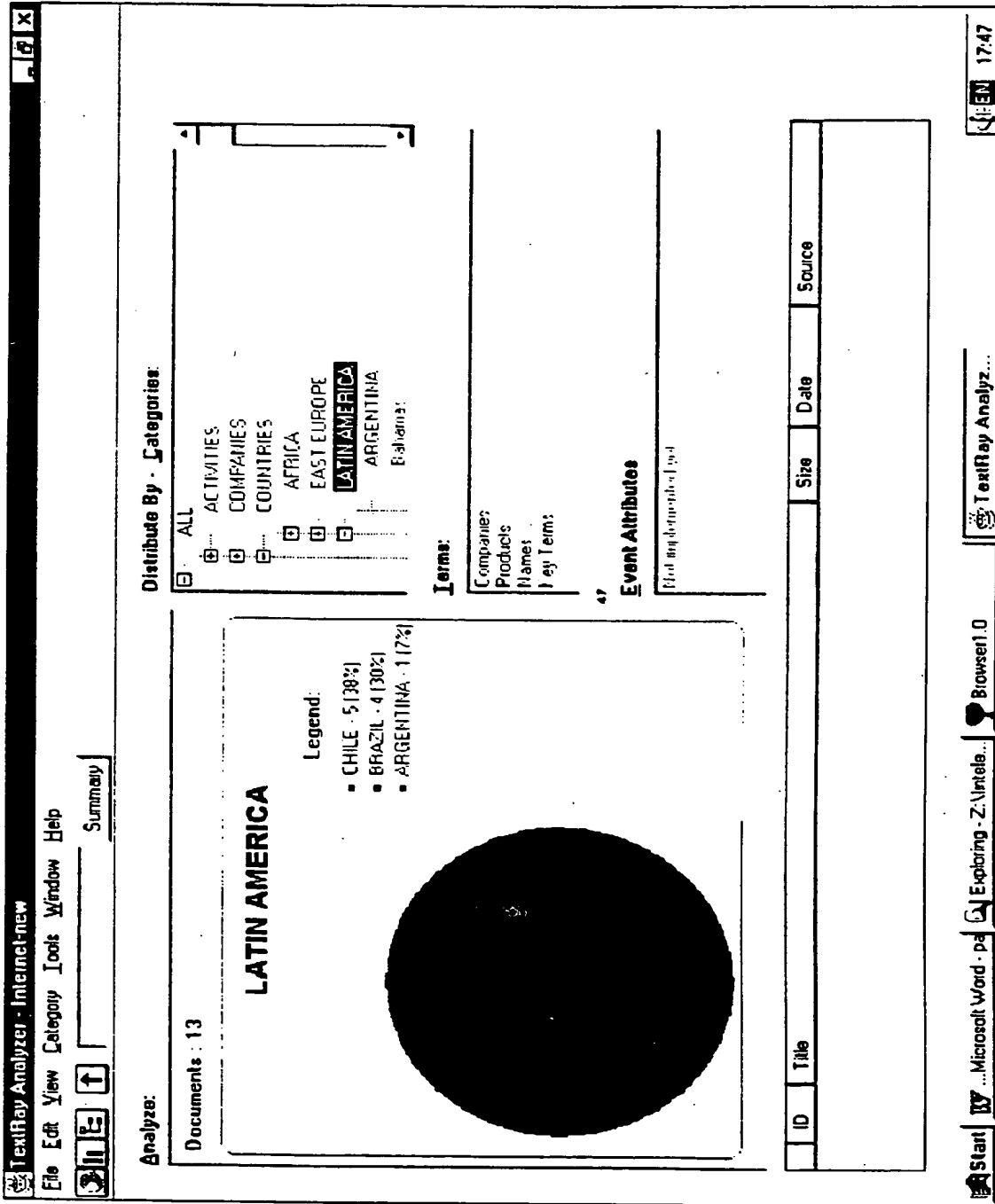


FIG.7

8/21

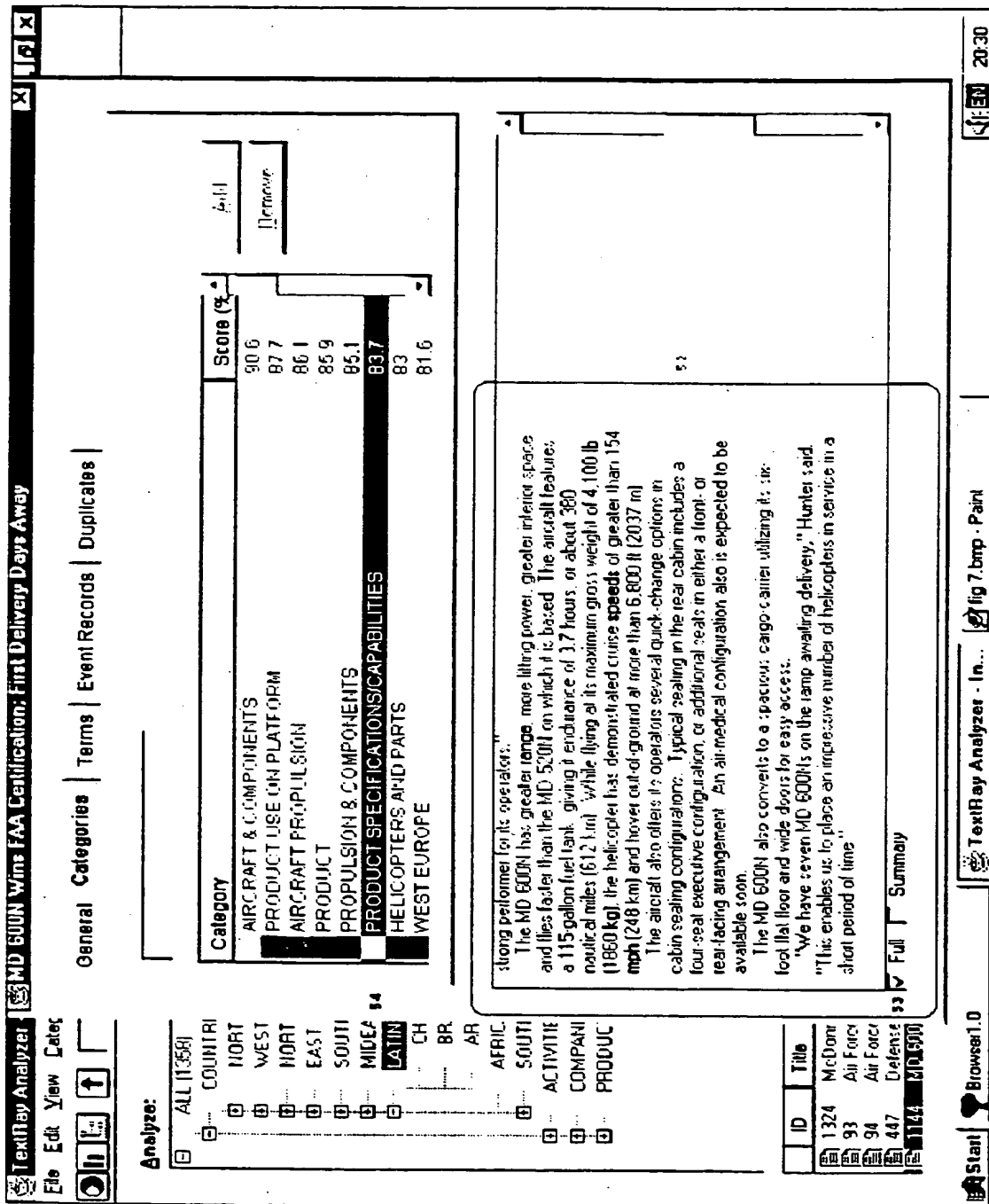


FIG.8

9/21

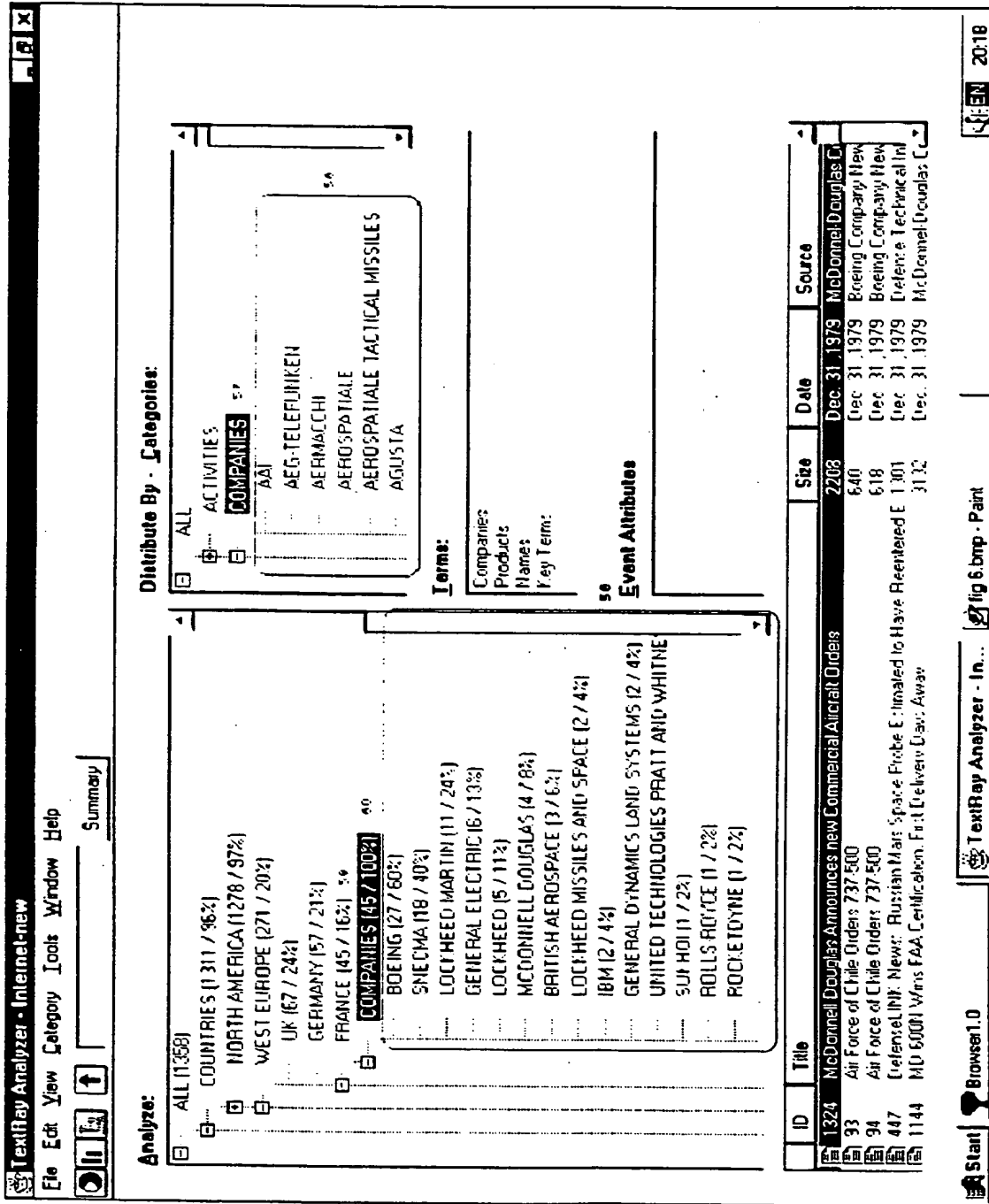


FIG.9

10/21

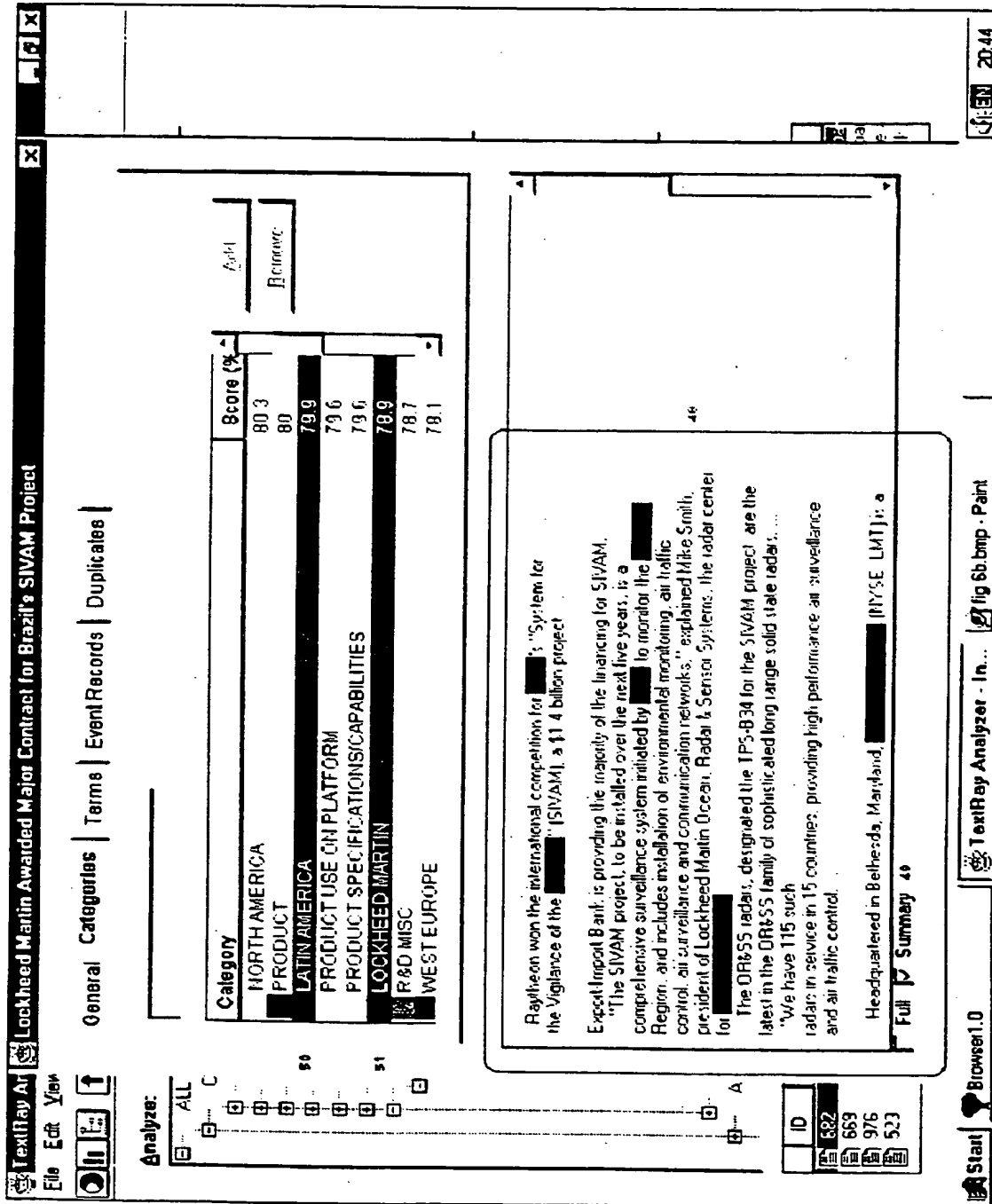


FIG.10

11/21

TextRay Analyzer - Internet-new

File Edit View Category Tools Window Help

MD 600N Wins FAA Certification; First Delivery Days Away

General Categories | Terms | Event Records | Duplicates

Analyze: ☐ ALL (1750) ☐ COUNTRIES (1311 / 96%) ☐ NORTH AMERICA (1278 / 97%) ☐ WEST EUROPE (271 / 20%) ☐ NORTHEAST ASIA (108 / 8%) ☐ EAST EUROPE (100 / 7%) ☐ SOUTHEAST ASIA & PACIFIC (55 / 4%) ☐ MIDEAST (53 / 4%) ☐ LATIN AMERICA (113 / 0%) ☐ CHILE (5 / 3%) ☐ BRAZIL (4 / 3%) ☐ ARGENTINA (1 / 7%) ☐ AFRICA (8 / 0%) ☐ SOUTH ASIA (7 / 0%) ☐ ACTIVITIES (1289 / 94%) ☐ COMPANIES (1160 / 85%) ☐ PRODUCTS (1102 / 81%)

02

Category	Score (%)
ARGENTINA	100
AIRCRAFT & COMPONENTS	90.6
PRODUCT USE ON PLATFORM	97.7
AIRCRAFT PROPULSION	86.1
PRODUCT	85.9
PROPULSION & COMPONENTS	85.1
PRODUCT SPECIFICATIONS/CAPABILITIES	83.7
HELICOPTERS AND PARTS	83

03

The customer kit includes deliveries scheduled in the United States, Switzerland, South Africa, Mexico, Argentina, Australia, Norway, Belgium, France, Finland, Japan, Hong Kong, Germany, the United Kingdom, Brazil and Venezuela.

"The combination of quiet, reliable performance and low operating costs is desired to make the helicopter a major factor in the light single market," said Eric Hunter, vice president of the company's light helicopter division.

"The helicopter has been thoroughly tested and found to be a top performer. It is a great addition to our product line," he said. "And, it will be a strong performer for its operators."

The MD 600N has greater range, more lifting power, greater interior space, and flies faster than the MD 520N on which it is based. The aircraft features a 115-gallon fuel tank, giving it endurance of 3.7 hours, or about 380 nautical miles (512 km). While flying at its maximum gross weight of 4,100 lb (1,860 kg), the helicopter has demonstrated cruise speeds of greater than 154 mph (248 km/h) and hover out-of-ground at more than 6,800 ft (2,073 m).

The aircraft also offers its operators several quick-change options in cabin seating configurations. Typical seating in the rear cabin includes a

Full Summary

Start | Notes - Mbr... | Exploring... | Find: Files n... | Microsoft W... | Browser 1.0 | TextRay... | 27 | Unfiled - Pa... | 2013

ID	Title
1144	MD 600N Wins FAA Certification; First Delivery

FIG.11

12/21

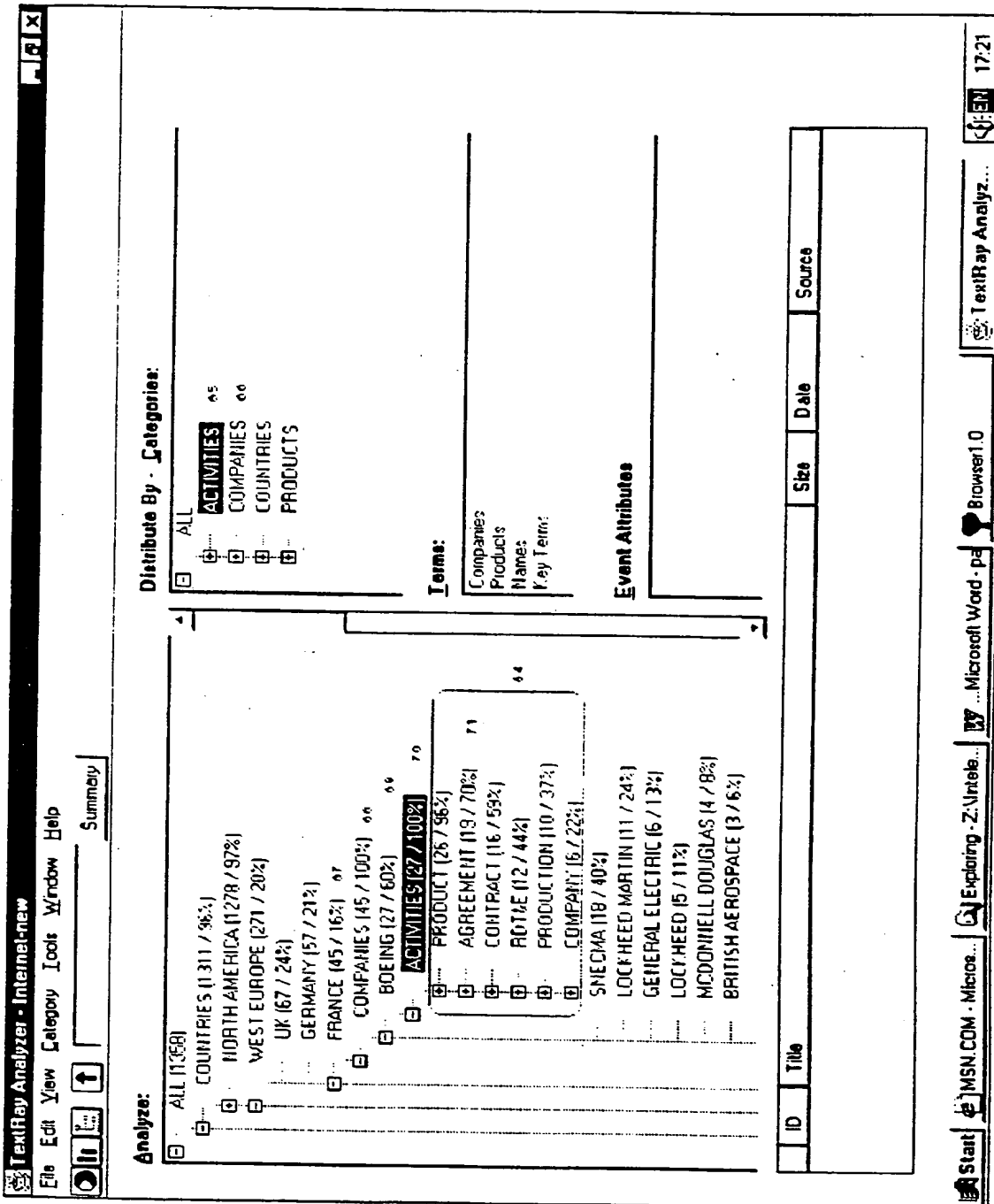


FIG. 12

13/21

TextRay Analyzer - Internet-new
 File Edit View Category Tools Window Help

Analyze:

Distribute By - Categories:

☐ ALL ☐ ACTIVITIES ☐ COMPANIES ☐ COUNTRIES ☐ PRODUCTS

Terms:

☐ Companies ☐ Products ☐ Names ☐ Key Terms

Event Attributes

Summary

ALL (1358)

☐ COUNTRIES (1311 / 96%)

☐ NORTH AMERICA (1278 / 97%)

☐ WEST EUROPE (271 / 20%)

UK (67 / 24%)

GERMANY (57 / 21%)

FRANCE (45 / 16%) 55

NETHERLANDS (24 / 8%)

ITALY (24 / 8%)

SWITZERLAND (23 / 8%)

SPAIN (16 / 5%)

FINLAND (14 / 5%)

BELGIUM (11 / 4%)

SWEDEN (3 / 3%)

NORWAY (4 / 1%)

GREECE (4 / 1%)

AUSTRIA (4 / 1%)

TURKEY (3 / 1%)

DENMARK (3 / 1%)

☐ NORTHEAST ASIA (108 / 8%)

☐ EAST EUROPE (100 / 7%)

☐ SOUTHEAST ASIA & PACIFIC (55 / 4%)

ID Title Size Date Source

208 Boeing 737-700 Generates More French Jobs Than Airbus A319 2078 Dec. 31, 1979 Boeing Company New

204 Breaking Down Barriers in Aerospace Industry Is Good For All of Us, Says Boei 4188 Dec. 31, 1979 Boeing Company New

331 DefenseLink: News: DOD SELECTS FOREIGN DEFENSE EQUIPMENT FOR T-21651 Dec. 31, 1979 Defense Link

142 Air France Orders 10 Boeing 777 Jetliners 758 Dec. 31, 1979 Boeing Company New

1315 McDonnell Douglas Announces Sales of 13 Helicopters 2060 Dec. 31, 1979 McDonnell Douglas Co

Start **Contact...** **Employ...** **Find: Fil...** **Micros** **fig 8.b...** **fig 6.b...** **fig 3.b...** **Browse...** **Test...** **18:45**

FIG.13

14/21

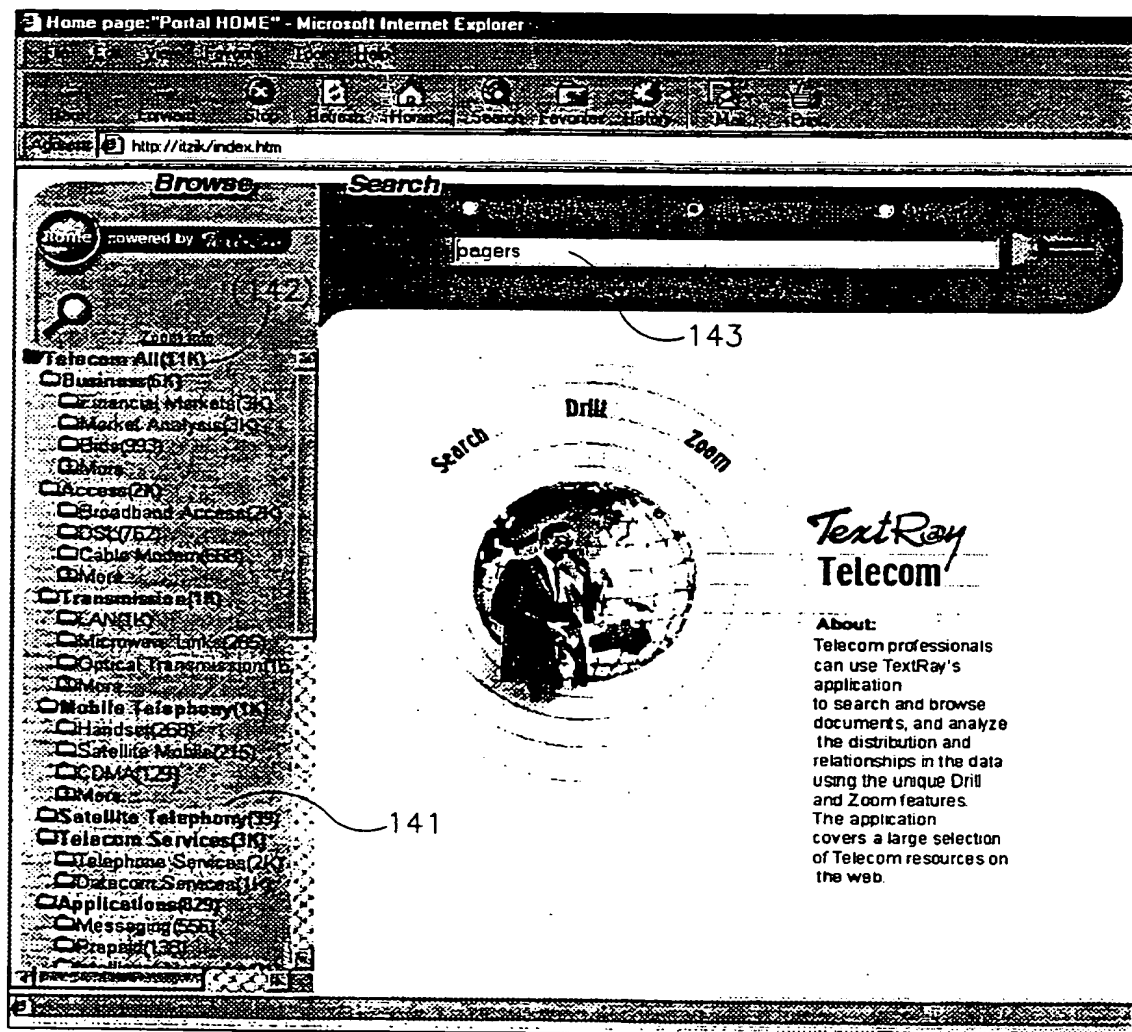
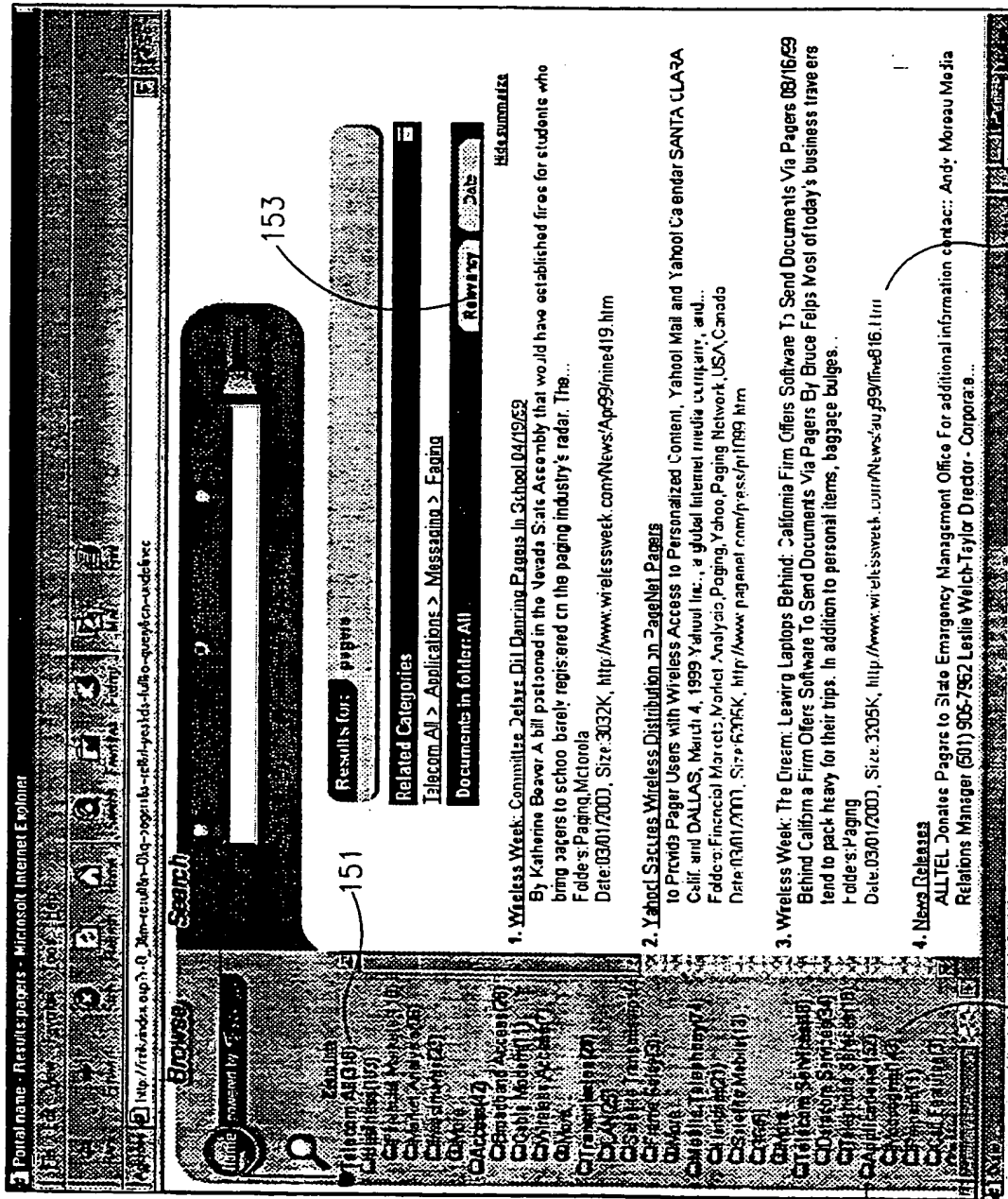


FIG.14

15/21



152

155

FIG.15

16/21

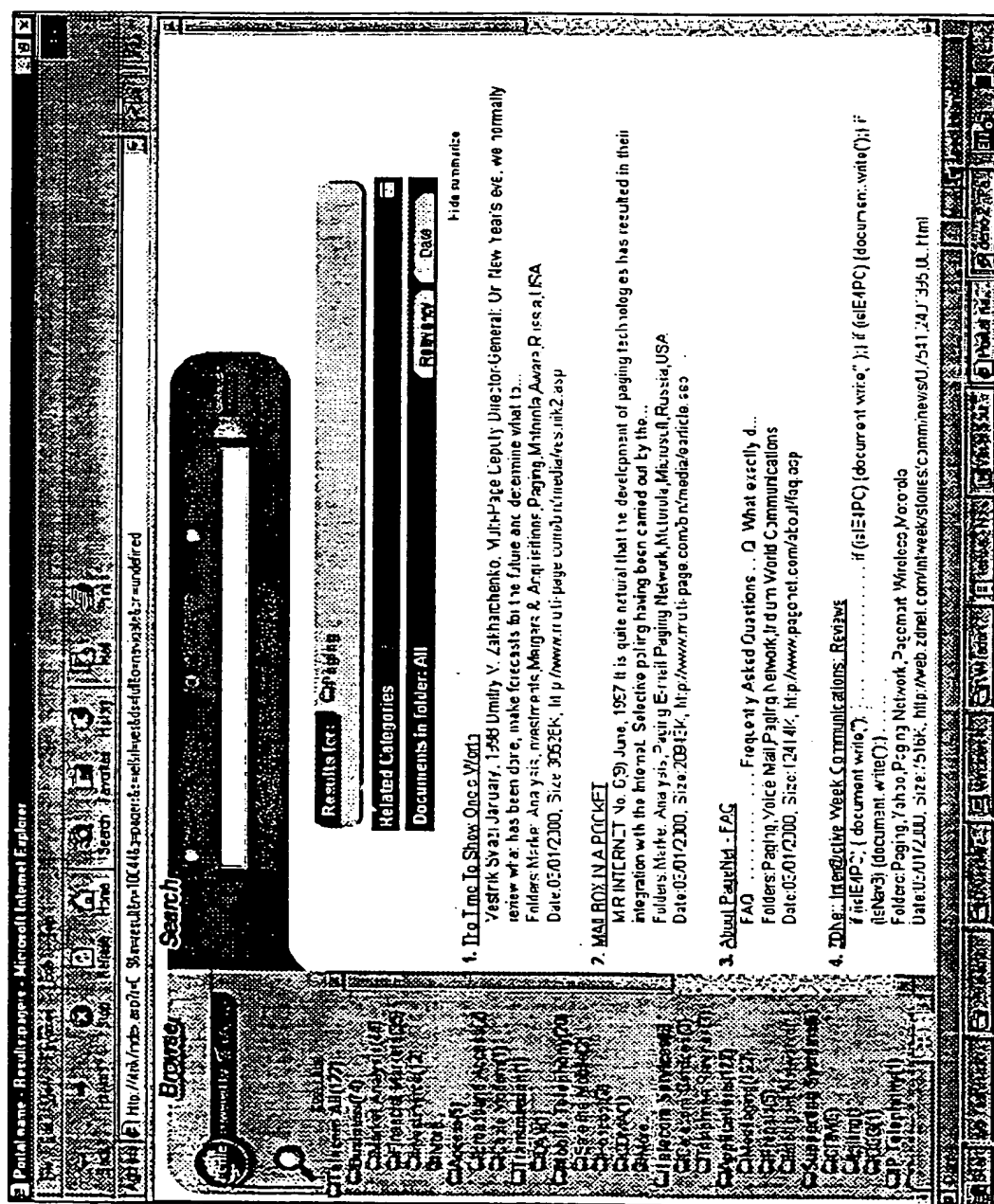


FIG.16

17/21

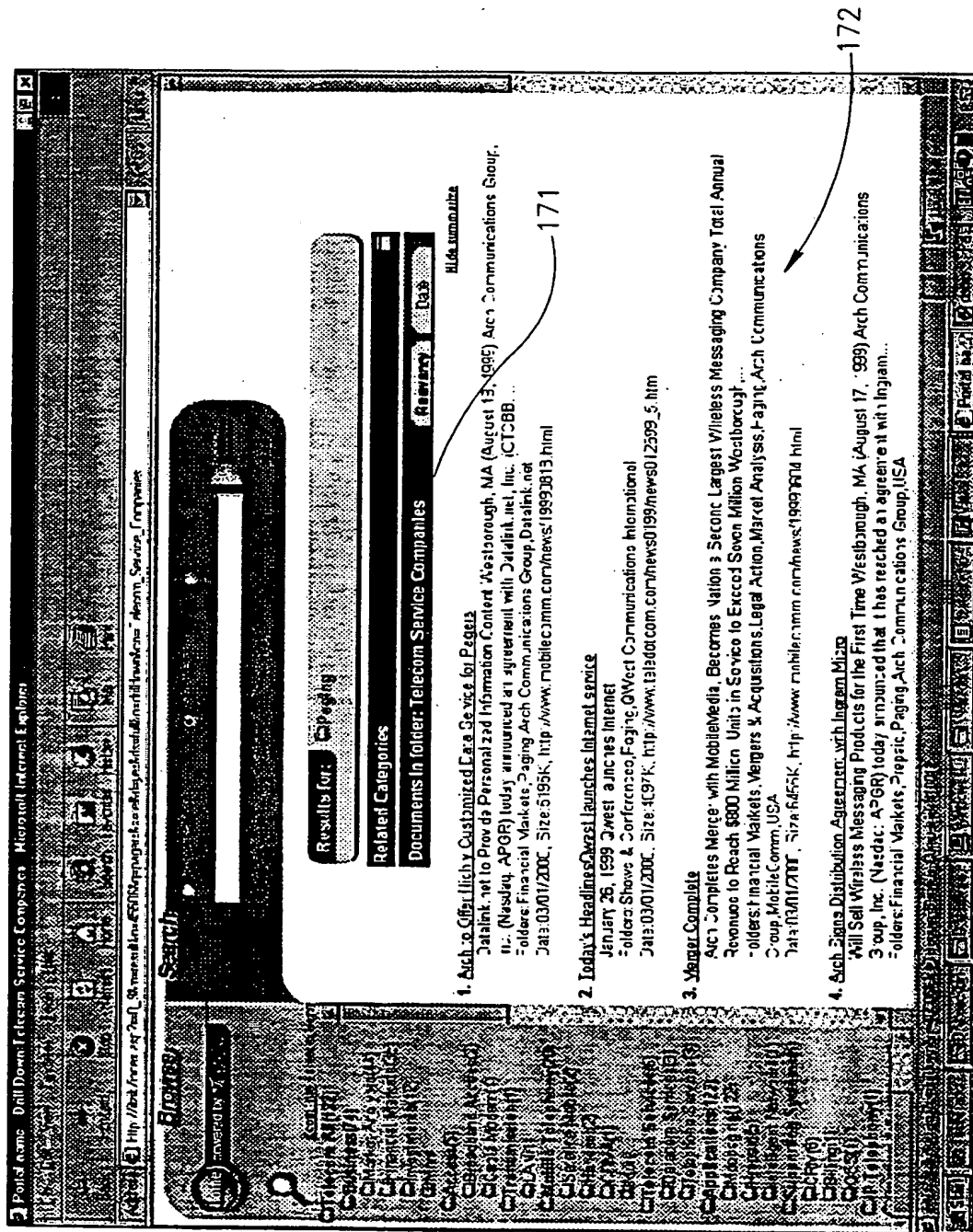


FIG.17

18/21

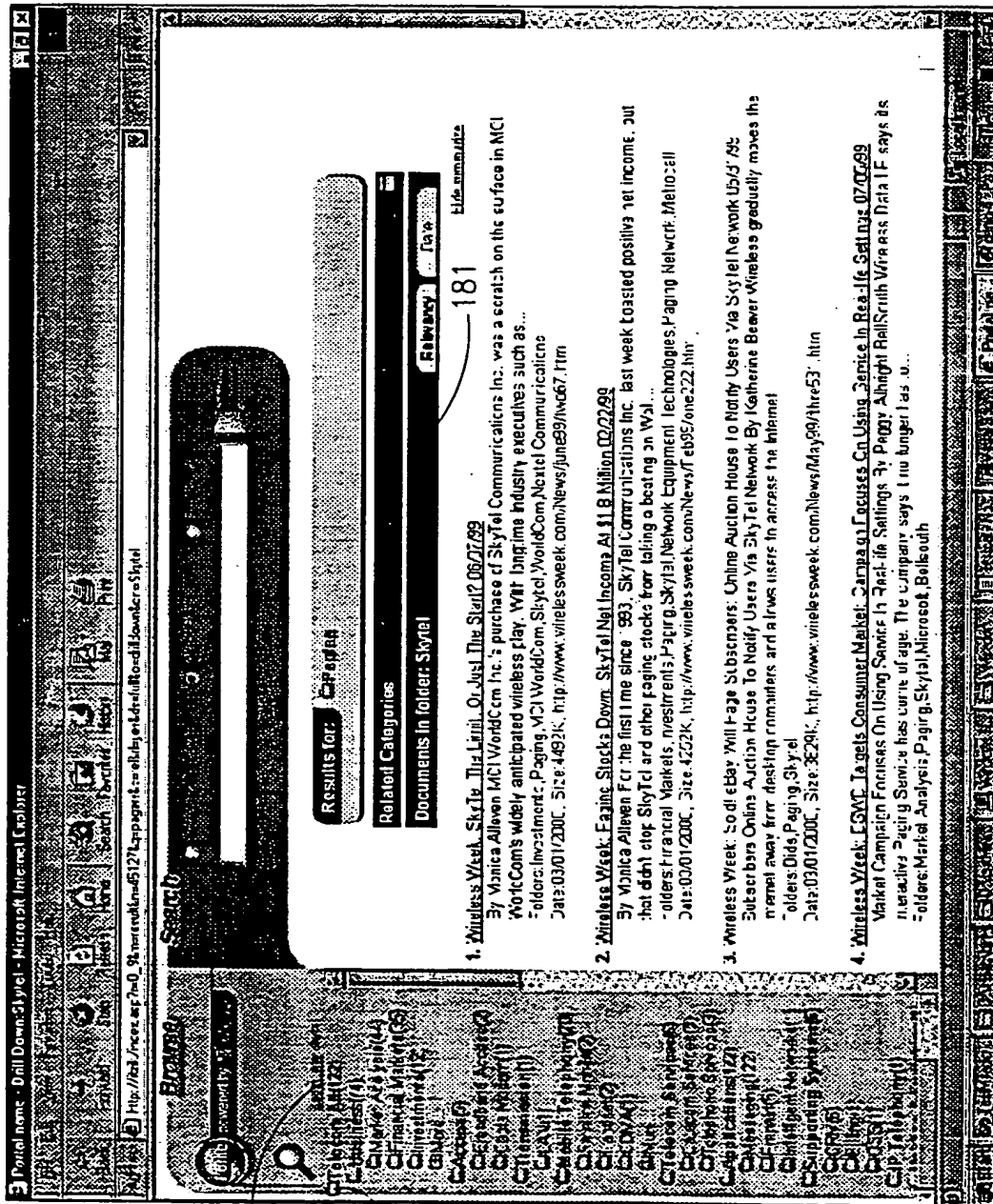


FIG.18

19/21

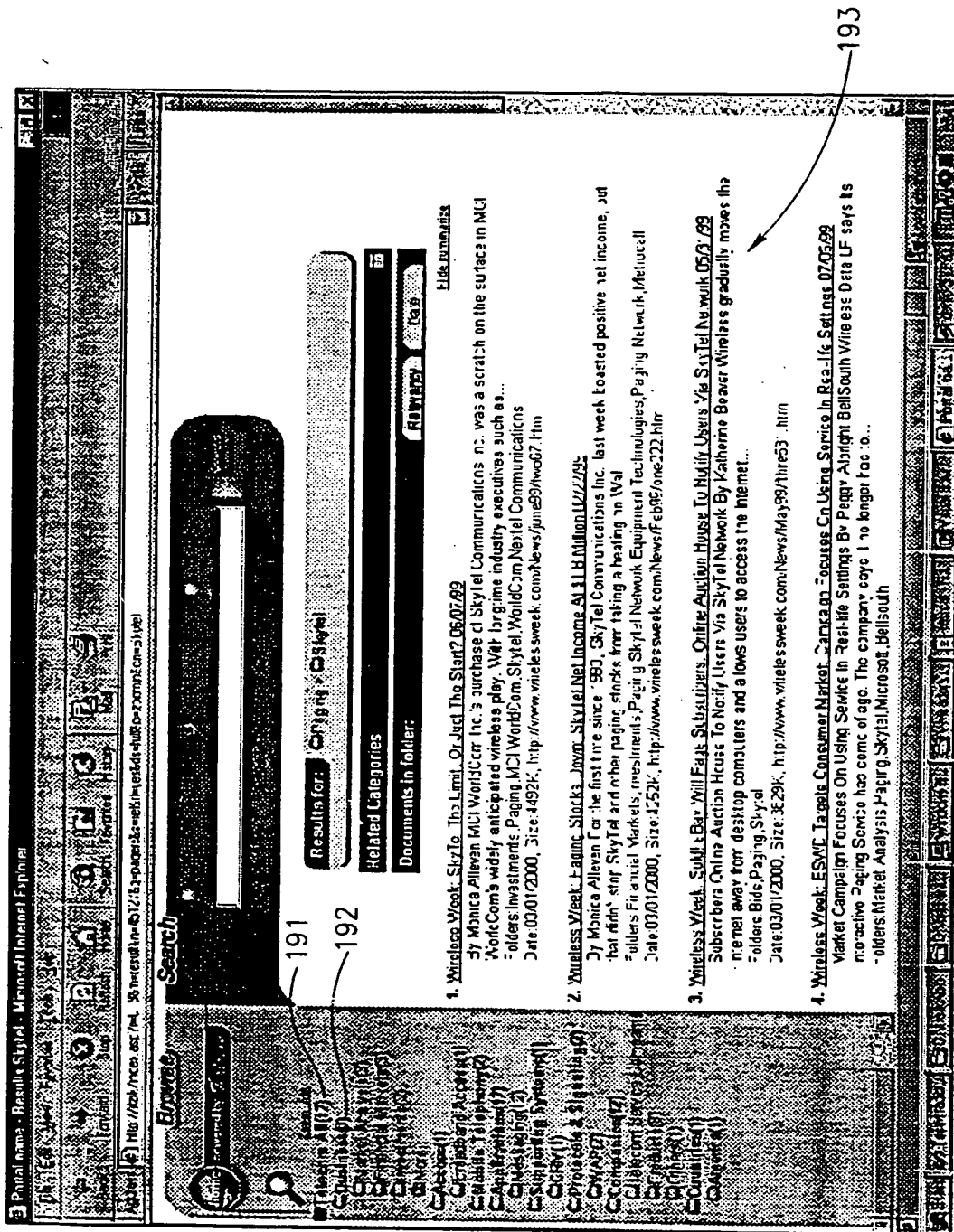


FIG.19

20/21

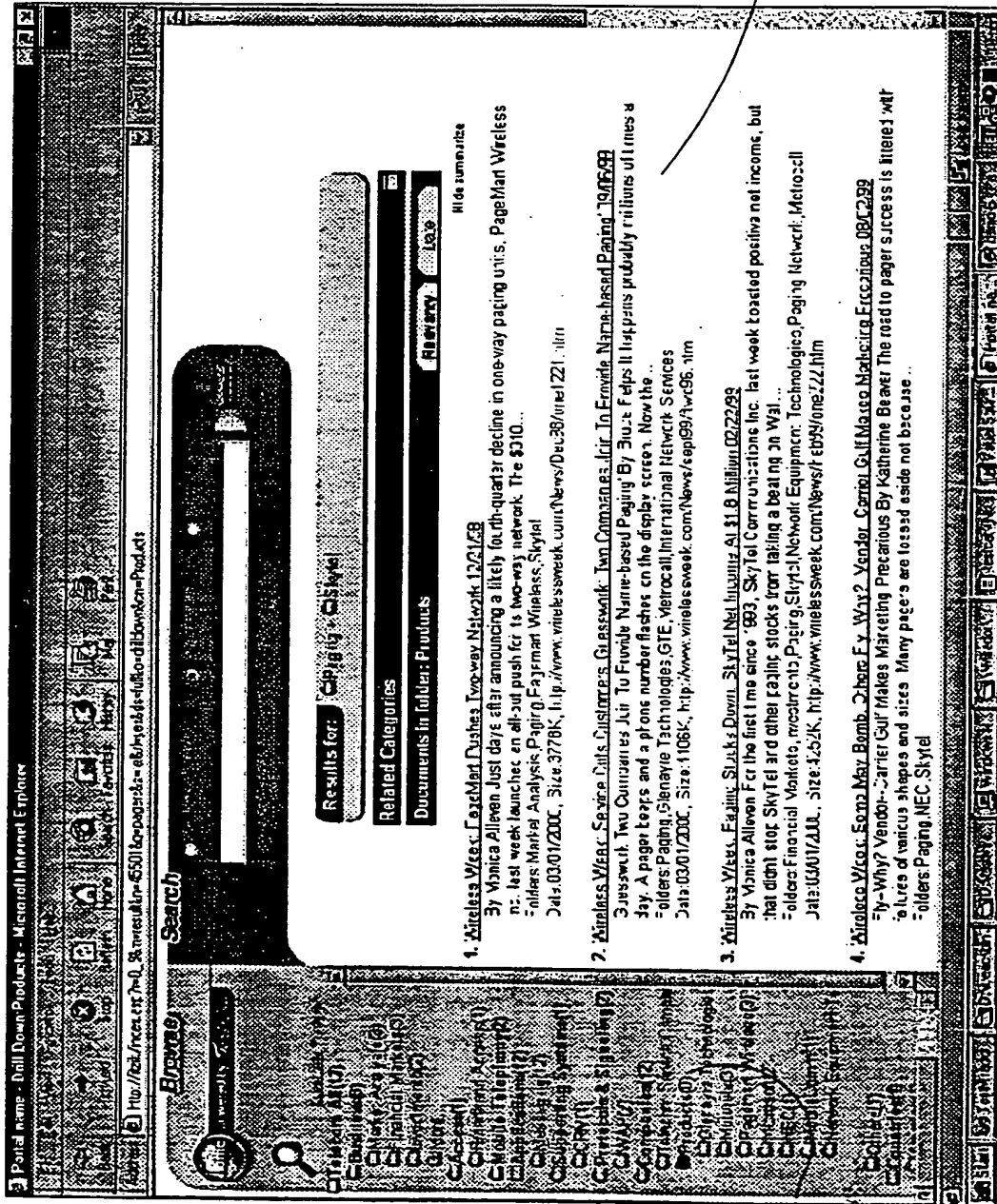


FIG.20

21/21

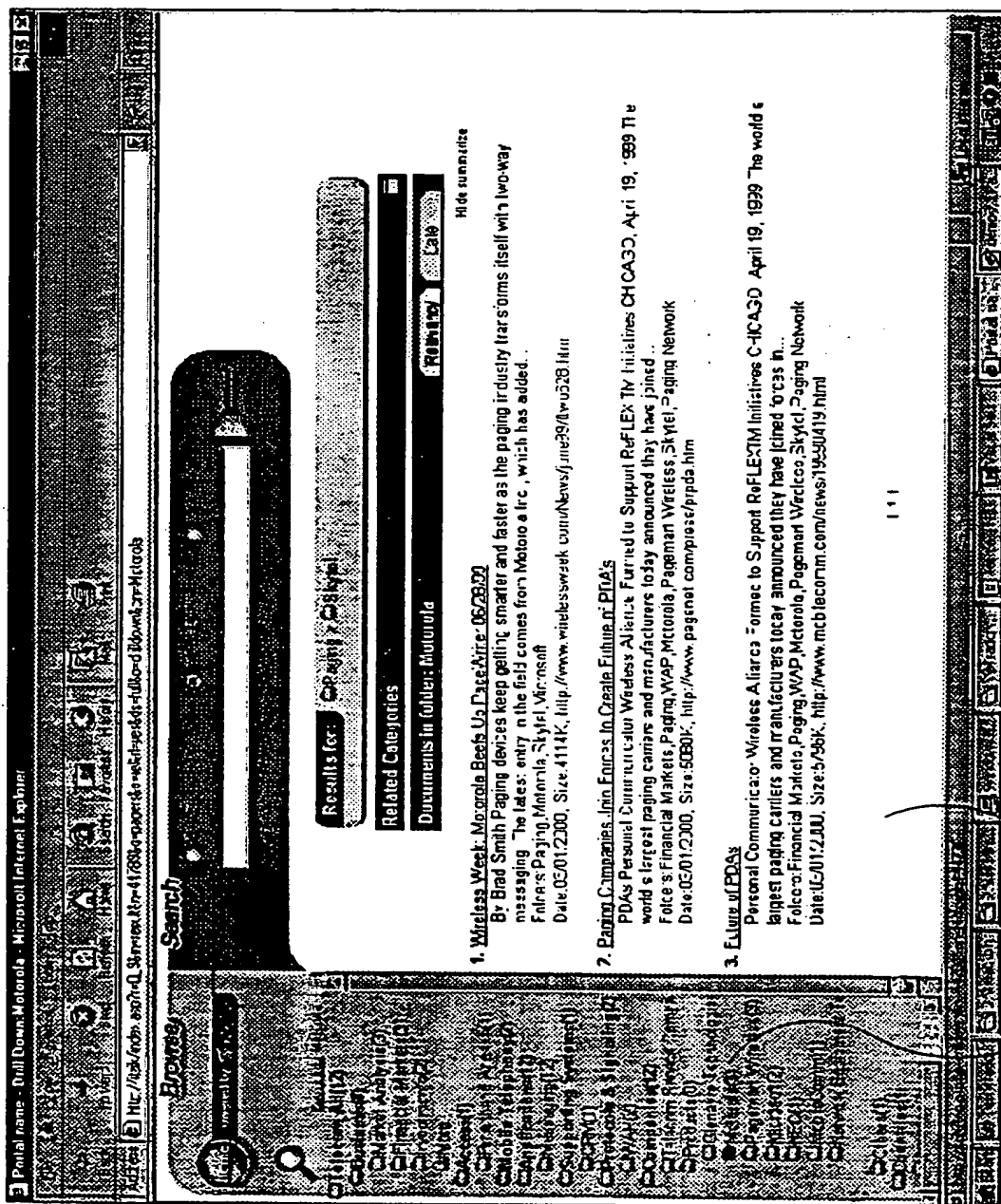


FIG. 21

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IL 00/00117

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, EPO-Internal, PAJ, INSPEC, COMPENDEX, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHING-CHI HSU ET AL: "Constructing personal digital library by multi-search and customized category" PROCEEDINGS TENTH IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE (CAT. NO.98CH36294), PROCEEDINGS OF 10TH INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE (ICTA'98), TAIPEI, TAIWAN, 10-12 NOV. 1998, pages 148-155, XP002141059 1998, Piscataway, NJ, USA, IEEE, USA ISBN: 0-7803-5214-9	1,2,5, 11,17, 18,24, 26-28,31
Y	page 149, left-hand column, line 19 -page 150, right-hand column, line 12; figures 2.1,2.2,2.3 page 151, right-hand column, line 48 -page 152, left-hand column, line 24; figure 3.2 -/--	12,13

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

26 June 2000

Date of mailing of the international search report

12/07/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Polzer, A

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IL 00/00117

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p style="text-align: center;">---</p> <p>CHAKRABARTI S ET AL: "Using taxonomy, discriminants, and signatures for navigating in text databases" PROCEEDINGS OF THE TWENTY-THIRD INTERNATIONAL CONFERENCE ON VERY LARGE DATABASES, PROCEEDINGS OF VLDB 97: 23RD INTERNATIONAL CONFERENCE ON VERY LARGE DATABASES, ATHENS, GREECE, 26-29 AUG. 1997, 'Online! pages 446-455, XP002141060 1997, San Francisco, CA, USA, Morgan Kaufmann Publishers, USA ISBN: 1-55860-470-7 Retrieved from the Internet: <URL:http://www.vldb.org/dblp/db/conf/vldb/ChakrabartiDAR97.html> 'retrieved on 2000-06-26!</p>	<p>19,21, 22,29,30</p>
Y	<p>abstract page 446, right-hand column, line 40 -page 447, left-hand column, line 22 page 447, right-hand column, line 39 -page 448, left-hand column, line 43</p> <p style="text-align: center;">-----</p>	<p>12,13</p>